

19 March 1981

US versus science education

The first Reagan budget, well advertised in advance as a device for getting the government off the taxpayer's back and the taxman out of his pocket, has more than lived up to expectations in its mean view of science education, at least as administered by the National Science Foundation. Ever since the heady post-sputnik days, when Washington was thronged with committees giving the Eisenhower Administration advice on how best to meet this unexpected challenge to American technology, the National Science Foundation has been spending a sizeable part of its budget (but still a modest annual sum) on the support of curriculum development in schools and colleges. It can boast of having done more good than foolish works. Zacharias's organization, the Physical Sciences Study Committee, may not have produced the ideal physics curriculum for the average high school, but it was a cheerful demonstration that distinguished academics were willing and able to help with the problems of teaching in the schools. And many young people (not only in the United States) learned a great deal of physics from the textbooks, readers and experiments that the committee devised. In retrospect, however, it may have been even more telling that the experiment proved infectious. Quite soon, in the early 1960s, an army of high-school teachers and a battalion of academics were caught up in the network of committees dedicated to the improvement of science teaching.

Inevitably, there have been excesses and mistakes. If the foundation was less confident in the development of curricula in the social sciences and the humanities than in the physical or biological sciences, it should not shoulder all the blame. Mere money is not sufficient to design a good curriculum. The essential is that good teachers and perceptive academics should be willing, and that school systems should help with administration and facilities. In any case, what matters is not that such and such a set of books should be written as an integrated package (and then, with luck, adopted in the Californian schools) but that teachers should have learned that continual change is necessary but also

possible. The \$13 million the foundation spent on these modest efforts last year will not go far to help the US Treasury to balance its books (or, rather, to live with a deficit of \$41,000 million). Cutting back on this part of the foundation's programme when the high schools are painfully aware of the need for new kinds of teaching seems either unwise or spiteful. Or is the Administration calculating that with Mr Walter Cronkite's new science programme advertised on CBS for the summer, the National Science Foundation can fade gracefully from the scene?

The budget request will also cut the foundation's modest support for students of science and engineering, mostly at the graduate level. The timing is entirely misplaced. Only a few months ago, the Carter Administration was agonizing about the problems of recruiting good people into engineering and especially engineering education. The problem cannot have disappeared since Inauguration Day. And in spite of the largely self-sustained character of American students — the envy of British university administrators (see below) — the United States government is probably still getting value for money from its spending on this good cause. Any Presidential Science Advisor would vouch for that, which may be one of the reasons why President Reagan appears reluctant to appoint such a person.

One of the temptations that afflict incoming administrations everywhere is that of abolishing at least some pre-existing institutions. The National Science Foundation's education programme is a natural candidate for this treatment. What it has been doing is what the market might otherwise have to do. It is also an administrative inconvenience that there should be an organization concerned with education entirely separate from the Department of Education. And for all their clubbiness, people interested in such things as curriculum development do not constitute a constituency of any importance. So why not abolish the institution? The irony, which other administrations have found to their cost, is how often it is necessary to recreate a substitute of some kind.

. . . . and UK versus universities

British universities, after too many months of somnolence, have taken fright. Last week, Sir Alec Merrison, Vice-Chancellor of the University of Bristol but speaking in his capacity as chairman of the Committee of Vice-Chancellors and Principals, accused the government of "a kind of madness". He and a posse of fellow vice-chancellors seem privately to have told the Secretary of State for Education and Science, Mr Mark Carlisle, much the same; by Merrison's account, they made no impression with their complaints that the British government is "profoundly misguided" in its intention to cut state support for the university system by 8.5 per cent in the next three years. Together with the switch to full economic fees for overseas students now under way, state support for the British university system will have fallen by at least 11 per cent, and perhaps by 15 per cent, by the academic year beginning in 1983. Either way, structural changes of some kind will be forced on the system. It is no wonder that people have taken to strong language.

Universities themselves must, however, shoulder some of the responsibility for what is now in prospect. At least since the summer of 1978 (under a different government), it has been clear that structural change was unavoidable. Even the now forgotten promise of "level funding" would not have allowed universities

to keep on recruiting young people to academic posts at a sufficient rate to stay intellectually alive, also competitive. The age distribution of people in tenured posts, skewed by the rapid recruitment in the 1960s of then young academics, ensures that in the 1980s the rate of retirement among British academics (now one per cent a year) will be less than a third of that needed to permit recruitment at a rate corresponding to a steady state. British universities have talked at length about the problem; none has found a solution. The government's decision, two years ago, that universities must collect larger tuition fees from overseas than from home students, whose first effects are becoming apparent only in this academic year, has had the effect of distributing penury at random. Again, most universities have put on hair shirts to wait for whatever the fates might bring. The outstanding exception is the London School of Economics, which has actually increased the number of its new overseas customers by the simple expedient of telling prospective students (and their governments) about its own virtues. The new cut goes deeper. The government intends simply to take cash out of the system, and there is no alternative source in sight. The damage that will be done would have been more easily contained if the university system had responded much more imaginatively to the previous



crises. There has been plenty of warning.

The government's behaviour is even less creditable, but consistent with its brief record. Overseas students' fees were increased by fiat, without consultation (and without much thought). The result is not merely the arbitrariness with which different universities are affected but the patchwork of petty anomalies and injustices with which university administrators have to struggle. The latest cut has been deviously introduced. Last December, when the government casually announced a cut of 3.5 per cent in next year's budget, nobody mentioned that there was worse to come. And last week the announcement of two further instalments (in 1982 and 1983) was tucked away in the Treasury's annual forecast of government expenditure. Was somebody hoping that nobody would notice? Or that the issue would be submerged in the generalized sense of grievance which this year's higher taxes have evoked? Whatever the truth, there seems to have been only the most cursory consultation with the universities about the consequences of the new cuts and no serious attempt to discover the least damaging ways of saving money. In short, there is no explicit policy for the universities. On all the evidence, there seems not even to be a secret policy. Ministries involved with higher education utter threats and then, several weeks later, deny them with reassuring promises. A few weeks ago, for example, the government had given the universities to understand that they should not try to solve their money problems by taking in extra students; last week's documents on spending promises that they can take as many as they can teach. Is it not time that Parliament forced the government to say what is in its mind? Have the universities no friends left at Westminster?

Ironically, the academic community is probably as keenly aware of the government's economic problems as any other in British society, with the possible exception of the bankers, themselves afflicted by a retrospective capital levy in this year's British budget. The government's central objective is to borrow less from the commercial markets, to which end spending less is important. The new cuts will yield £80 million a year and the increased fees for overseas students nearly as much. However the "saving" is shared out among the British universities, the consequences will be bad for the quality of British university education. The government's expenditure plan says laconically (in a one-sentence paragraph) that "the plans assume a significant tightening of staffing standards", which is a euphemism for a further decline in the ratio of teachers to students. The plan also says that the universities collectively will admit fewer students in the years ahead. The plain implication is that the system will have to shed teachers. That, in the last resort, is why the vice-chancellors are up in arms.

So how should the universities respond? As they should have done in the previous crises of the past three years, by organizing themselves for flexibility. With a few exceptions, the objectives of most university institutions in Britain are similar — the pursuit of the highest academic standards in the teaching of undergraduates and of excellence in research. For most universities and almost all academics, the ideal is Oxbridge. And there is very little encouragement, in the way the government subvention is disbursed on the recommendations of the University Grants Committee, for them to aim at distinctive goals. The result is that too many British universities are trying to do the same things with such inappropriate resources that they are bound to fail.

Would it not, however, be preferable that they should aim at excellence in what they already do well? The system includes half a dozen institutions which are what would be called liberal arts colleges in the United States but which nevertheless waste endless time and money on graduate education in the traditional pattern without recognizing that their graduate schools will always be too small to be viable. It would be better for the institutions concerned, their students and the system as a whole, if this were recognized. The strong civic universities in Britain, foundations of the late nineteenth century, are another distinctive category of universities. Not so long ago, they drew a great deal of strength (and substantial amounts of cash) from the cities in which they were sited. Although local loyalties are still in many places strong,

they are nowhere as strong as they used to be when it was no shame for a young person to become a student at the nearest university. In the serious crisis about to inflict itself on them, the civic universities will regret the way that the old loyalties have been eroded. Some of them may even find themselves in sneaking sympathy with the plan which the government is said to be hatching for replacing the present system of maintenance grants for students (over and above tuition fees) with a smaller grant and the offer of a loan. There could hardly be a better device for channelling students towards the nearest university, while stronger local links would give these universities opportunities to increase their present interest in the retraining of local adults — professionals, technicians, even craftsmen.

The benefits of reorganization for diversity are several, not least the direct economies there would be once the universities stopped aping each other. In such a pattern, the system could hope to match courses to students more expertly than at present, while a clear knowledge of their distinctive roles would make it possible for individual institutions to tell a more convincing tale when asked (as they will be frequently in the years ahead) why they continue to exist. But how is such a pattern (or any other pattern) to be brought about? The conventional wisdom among academics is that universities are so incapable of making long-term policy, and of imposing change upon themselves, that change must be forced from outside, either by the University Grants Committee or by some seemingly impartial committee made up of academics from elsewhere. This is too depressed a view.

Although the occasions in the past when individual universities have changed radically have been times of plenty, there is no reason why the penury now in prospect should not also be a spur to change. Sheer self-interest dictates that it should be, but the University Grants Committee could do much to help. Institutions willing to take painful steps — closing, say, a particular department, or all graduate schools — need and deserve compensation. The snag, in present circumstances, is that an institution making substantial savings in its internal operations cannot be sure that it will benefit in the long run. The University Grants Committee can always in its wisdom recommend that the funds should be spent elsewhere in the system. So would it not make sense, for the duration of whatever transitional period lies ahead, that institutions making voluntary sacrifices should benefit from some kind of salve for their wounds? The University Grants Committee, more used to sharing resources according to its own hidden definitions of equity, is not used to the encouragement of entrepreneurship. But even the University Grants Committee will have to change before this crisis is over.

So too will many other institutions. Even if the outcome of the next few years is a concentration of graduate education in a subset of the universities, it would be intolerable if academics and graduating students from elsewhere were denied opportunities for research. This would require that sources of research support, the research councils in particular, should change. Then, given the need for compatibility within a mobile and diversified system, most universities would find it necessary to restructure their undergraduate curricula. Many, no doubt, would rediscover the virtues of the four-year undergraduate courses, of less specialized (if still competitive) entry requirements and of less narrow courses of study. In other words, there may be benefits in the present adversity.

In the wake of last week's shock to British universities, many despondent academics will consider this cheerful outcome a delusion. Is it not more likely, they will ask, that the University Grants Committee will recommend that misery should be evenly distributed, with the result that all universities become less good? That, certainly, is possible. It is, however, unlikely that all universities within the system will be equally passive. When they have recovered their breath, the strongest among the universities will be thinking of how to secure for themselves a greater measure of independence than they have enjoyed for years. In due course, all those that survive will follow suit. For what the latest crisis in the British system shows is that universities, jealous of their claims to autonomy, are safest when they are also independent.

Kiss of life for Clinch River breeder

Reagan ditches Carter policy on proliferation

Washington

Despite the objections of his economic advisers, President Ronald Reagan has decided to pay off some of his political debts by letting construction work begin on the liquid metal fast-breeder reactor at Clinch River in Tennessee, which critics claim is both economically unnecessary and technically out of date. However, Mr Reagan is holding back from providing further federal support for the reprocessing of spent nuclear fuel until it becomes clear whether the private sector is prepared to take on this responsibility.

Both the fast breeder programme and commercial spent-fuel reprocessing had been held in check by President Carter on the grounds that they contributed to the growth of the "plutonium economy" and thus raised the risk of the proliferation of nuclear weapons. But both decisions have come under strong attack from the nuclear industry as an unnecessary restriction on the growth of nuclear power, an attitude known to be held by the new Energy Secretary.

In his budget request to Congress, the full details of which were released last week, President Reagan is requesting a total of \$737 million for support of fast-breeder research in the fiscal year 1982. A substantial proportion of this increase will go towards the start of construction work on the Clinch River reactor as a "demonstration project". Development funds have already paid for the design and much of the component fabrication, but construction had previously been held up by President Carter's request to the Nuclear Regulatory Commission not to issue a construction licence.

The Clinch River reactor has a high symbolic value for the US nuclear industry. Many used it to focus their opposition to Mr Carter's general non-proliferation policies. Others argued against the Carter Administration's conclusion that the declining rate of increase in the demand for electricity, and a short-term surplus in domestic uranium supplies, reduced the need for an intensive breeder development programme.

Environmentalists now point out that they have a new — and perhaps surprising — ally, Mr David Stockman, director of the Office of Management and Budget and the chief architect of Mr Reagan's austerity budget. Using the same arguments put forward by the Carter Administration, Mr Stockman circulated a letter to fellow con-

gressmen in 1977 describing plans to subsidize the construction of the Clinch River reactor as "totally incompatible with the free market approach to energy policy".

Critics of fast-breeder reactors, such as the Natural Resources Defense Council, claim that in order to find the money for increased spending on nuclear programmes in general, including the liquid metal fast-breeder reactor, the Department of Energy is intending to cut back significantly in research and development on alternative energy sources. It is proposed to cut the solar energy research budget, for example, from \$625 million in 1981 to \$241 million in 1982, and research and development in conservation from \$558 to \$195 million, while research on fission reactors would increase from \$1,166 to \$1,247 million.

Even the Office of Management and Budget is said to have argued that there is

an inconsistency in cutting off support for "demonstration" projects in solar and fossil energy, while supporting just such an approach for fast breeders; Mr Stockman is said to have put up a fierce battle against Clinch River funding, although he was less opposed to the increased budget request for design work on a larger, more advanced fast breeder.

In a report published by the subcommittee last week, a former member who left Congress last autumn, Republican congressman John Wydler of New York, reported on a visit to the British fast breeder programme, and used the British (and French) experience to bolster his argument that completion of Clinch River is the logical next step for the United States.

Mr Wydler also reported that British scientists were attracted to the idea of cooperation with the United States, a move

British academics up in arms

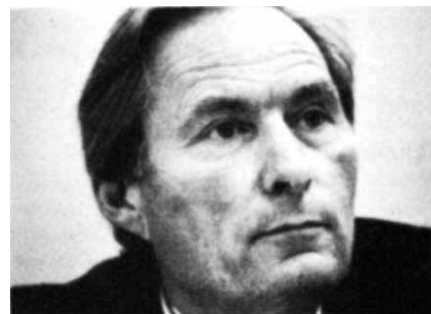
British universities were in a state of shock last week after the government's announcement of its spending plans for the next three years. Spending on home students is to be cut by 8 per cent in real terms by 1983–84. The Committee of Vice-Chancellors and Principals estimates that the total reduction in university income could be as high as 15 per cent when the shortfall in overseas students, expected after the government's decision to charge them economic fees, is taken into account. Sir Alec Merrison, chairman of the Committee of Vice-Chancellors, last week described the expenditure plan as madness. He said he was not aware that the government was reducing the universities' grant for any reason other than saving public money.

The cuts will undoubtedly force the universities to review the services they can offer, which is a fact acknowledged by the government in its white paper on expenditure published on budget day, 10 March. The vice-chancellors and the University Grants Committee seem to have persuaded the Secretary of State for Education that money should be set aside for the scaling down operation. But, instead of making more money available, as the universities must have hoped, the government announced on 13 March that £20 million should be taken out of the 1981–82 recurrent grant — at £979 million already 3 per cent less than for 1980–81 — for the purpose.

The Committee of Vice-Chancellors estimates that making necessary economies could be more expensive than maintaining the universities at their present level. The problem is that after several years of static budgets, further economies cannot be made without reducing staff. Salaries and related costs at present consume 72 per cent of the recurrent grant, more than half of

that going on academic pay. Redundancies, which would break the contracts of employment of tenured academics (approximately 80 per cent hold tenure), would involve the universities in costly litigation and large redundancy payments. And the committee's estimate that one in eight jobs will have to go would certainly mean the closure of departments and possibly entire universities. Nobody yet knows whether the £20 million set aside will be enough to help meet these costs.

The University Grants Committee will begin dividing what money there is between institutions in May. It is not thought to



"Madness" says Merrison

favour cutting any university's grant completely, but neither does it want to spread the money too thinly. Possible alternatives are to close selected departments — a new move for the committee, which normally leaves individual universities to run their own affairs — or to restructure the system, leaving some universities to develop as centres of excellence and turning others into something akin to American liberal arts colleges. The trouble is that the committee has never had to be so *dirigiste* in the past and has not formed a proper mechanism to take the necessary decisions.

Judy Redfearn

which had previously been endorsed by the committee, particularly because of problems with leakage in the existing British fast reactor steam generator components. Mr Wydler also said, however, that UK Atomic Energy Authority officials felt that cooperative agreements with the United States presented unique difficulties because of the Freedom of Information Act and the Atomic Energy Act requiring the disclosure of proprietary business information.

In contrast with his defeat in the fast breeder debate, Mr Stockman seems to have won on the other point of dispute with nuclear advocates in the Department of Energy, namely renewed federal support for spent fuel reprocessing from commercial reactors. The nuclear industry has been pushing for the reopening of the Barnwell reprocessing plant in South Carolina, which was denied funds by the Carter Administration; no money for commercial reprocessing is, however, included in the budget request, although there are funds for the West Valley Demonstration Project.

Backing up its general stance in favour of reprocessing, however, the Reagan Administration has decided to discontinue the spent-fuel programme intended to provide storage facilities away from power station sites. This programme is being refocused within the Department of Energy to concentrate on the development of alternative spent-fuel storage technologies which can be used by utilities at existing reactor sites.

Although the Administration has set its face against subsidies for commercial reprocessing, it has in the past few days become known that plans to dismantle at least this aspect of the Carter anti-proliferation policy are being worked out.

David Dickson

German big science

Win some, lose some

Germany should spend an extra DM 800 million (£170 million) on big science projects over the next ten years, recommends the committee set up a year ago by the previous minister of science, Volker Hauff — and another DM 425 million of projects should be accounted for under the current budget.

The "big science" committee, chaired by Professor Klaus Pinkau (who was recently appointed director of the Max Planck Institut für Plasmaphysik at Garching) reported last week to Minister Andreas von Bülow on a shopping list of ten major projects (*Nature* 3 July 1980, p.8) ranging from high-energy physics to geology, which, if they were all undertaken now, would have cost Germany at least DM 2,015 million (430 million). On being given the task, Pinkau said it would be a "work of art" to steer the right course between scientific worthiness and the government's

ability to pay — for the committee was given no financial guidelines. In the interim, the German economy has weakened, and the Pinkau committee felt obliged to take cost strongly into account.

However, some sleight of hand reduces the bill. The committee recommends that Germany support LEP, the large electron-positron collider which physicists would like to see as the next major facility at CERN, the European subnuclear physics laboratory in Geneva. LEP will cost Germany DM 350 million; but as it is likely to be built within CERN's current budget and staff, the committee reckons it will mean no extra expenditure. Similarly, a DM 75 million replacement for Meteor, Germany's principal and ageing oceanographic research vessel, can be paid for within the current geosciences budget.

Of the other projects, the committee says the refitting and expansion of the BER II research reactor at the Hahn-Meitner Institut, Berlin, costing DM 47 million, should be undertaken immediately; so should the construction of the DM 33 million SUSE, a 250 MeV superconducting cyclotron for heavy ion physics proposed by the Technical University of Munich — provided the university can raise part-finance from the regional government of Bavaria, and make the accelerator a national facility. An equivalent facility proposed by the Jülich nuclear laboratory near Cologne was rejected (it cost more).

The committee puts some of the most expensive projects on ice. HERA, a high energy proton-electron collider for the DESY subnuclear physics laboratory at Hamburg which would cost DM 600 million (in two parts) is recommended "in principle", but a decision should await developments in superconducting technology; and the question of a relativistic heavy ion accelerator for GSI Darmstadt (DM 190 million) is to "remain open".

The latter raises an interesting question in physics: will relativistic heavy ion collisions lead to entirely new states of nuclear matter (as in neutron stars) as some nuclear physicists predict, or will it just lead to a mass of high energy but otherwise familiar fragments? The feeling in Germany is that some low cost test of this question should be made first, for example, by feeding heavy ions into the intersecting storage rings at CERN. But if such experiments were done, German thinking goes, it should be *outside* CERN's budget or there will be no money for LEP.

On this issue, on HERA, and on others such as the proposed spallation neutron source (where Jülich and Karlsruhe have competing proposals) and the deep drilling programme of the Deutsches Forschungsgemeinschaft, Germany will also be looking for increased international collaboration. The European synchrotron radiation source, however, proposed by the European Science Foundation, is said to be "not of high priority". **Robert Walgate**

Madrid conference

Detente denied

Western leaders have failed to react "realistically" to Mr Brezhnev's latest proposals for detente, a *Pravda* editorial complained last week. The proposals, put forward at the party congress last month, covered three main issues: the extension of "confidence building measures" (CBMs) to all Soviet territory west of the Urals, and the possibility of introducing similar measures in the Far East, special negotiations on the Persian Gulf and Afghanistan, and a moratorium on medium-range nuclear weapons in Europe.

The party congress coincided with the closing weeks of the Madrid review conference on implementation of the Helsinki Final Act. CBMs, as specified in the Final Act, cover notification of major military manoeuvres (more than 25,000 troops), exchange of observers for manoeuvres and also notification of major troop movements, though this last clause is left to the discretion of the participating states. Notification of manoeuvres, moreover, in the case of a participating state whose territory extends beyond Europe, is mandatory only for an area within 250 km of its European frontier. The suggestions that the Helsinki clauses on CBMs should extend as far as the Urals, came, initially, from the French delegates to Madrid, and were not greeted with enthusiasm by the Soviet delegation, until, on 23 February, Mr Brezhnev made them his own. Not surprisingly, Mr Brezhnev suggested some kind of reciprocity from the West. Since the whole of Europe falls within the CBM zone, this would presumably have to be of a qualitative nature, such as the discretionary notification of smaller-scale troop movements, or, judging from hints in the Soviet media, curtailing "provocative" broadcasts to the Soviet bloc.

Afghanistan, too, was an issue which the Soviet delegation at Madrid at first tried to block as "irrelevant" (because the Helsinki Final Act referred to security and cooperation in Europe). Mr Brezhnev's call for a moratorium on nuclear arms seems to reiterate the Polish proposal at Madrid for a European disarmament conference in autumn 1981. Unfortunately, none of Mr Brezhnev's peace proposals touched on the sensitive issue of declaration and verification of troop and armaments strength — the point on which all similar negotiations have broken down for the past 35 years. His suggestion of an "authoritative international committee" whose members "might include the most prominent scientists" of the countries concerned, which would "demonstrate the vital necessity of averting a nuclear catastrophe" has, however, been passed on to the international Pugwash council as a matter of urgency.

On the other aspects of detente, the

Moscow seminar threat

The Moscow Sunday seminars for Jewish refusenik scientists, which have been held (official harassment permitting) since 1973, now face a new threat. Last Friday, Dr Irina Brailovskaya was informed by the deputy chairman of the local municipal council that if she continued to host the seminars, she and her whole family would be liable to internal exile.

This official told Dr Brailovskaya that there had been complaints from the neighbours about the holding of the seminars. "We do not object to scientific meetings", he explained, "but the devil only knows what is going on here!"

The seminars have only recently resumed after a KGB clamp-down last autumn. Following the arrest of Dr Viktor Brailovskii, on 13 November, for several weeks the apartment was surrounded on Sundays by security police who turned away intending participants. Early in the new year, the seminars resumed on Saturday evenings, and then, a few weeks ago, reverted to the traditional Sundays.

In a recent message to Western scientists, Dr Brailovskaya said that her husband has been held in custody beyond the legal limit (three months) within which he should have been either charged or released. He is, she said, in a poor state of health, suffering from chronic hepatitis and a blockage of the gall duct.

auspices seem good for renewed East-West cooperation. President Reagan's assistant secretary designate on human rights, Dr Ernest Lefever, is reported to favour keeping such matters distinct from international politics. The several thousand Western scientists who signed personal moratoria pledges bound themselves to limit their contacts with the Soviet Union only until the end of the Madrid conference. However, at the meeting of the United Nations human rights commission in Geneva last week, a motion from the Canadian delegation was dropped from the agenda.

Vera Rich

Science education

Tap turned off

Washington

The Reagan Administration is proposing to eliminate all support at US universities for undergraduate and graduate science students at present provided by the National Science Foundation.

The cuts are part of a plan to phase out the foundation's responsibilities for science and engineering education, justified by the claim that these are of lower priority than the need to support basic research. But the proposed cuts may have been partially instigated by conservatives upset by some of the foundation's educational activities.

The cuts would eliminate several programmes introduced over the past few years, primarily at the direction of Congress. These include support for women, minorities and "talented individuals" in science, strengthening science teaching in middle and high schools, and efforts to improve public understanding of science.

The new Administration argues that such efforts are of low priority, and that they are too broadly spread and narrowly focused to have any significant effect. Science teaching has, however, always been treated better by Congress than by either the National Science Foundation itself or the Office of Management and Budget; and there remains a good chance that some of the money will be put back during the congressional review process.

In response to the report last year by the foundation and the Department of Education which drew attention to the growing "scientific illiteracy" among the US public and to concern over the lack of adequate technical graduates in fields related to energy and military technology, President Carter had proposed in January that the 1982 budget for the foundation should include an allocation of \$119.9 million for science and engineering education, a significant increase over the \$80.7 million allocated in the current year.

In contrast, the new Administration suggests that the current budget be reduced to \$64.7 million, and that support for science and engineering education next year be reduced to \$9.9 million. All of this sum would be used to meet existing commitments to students receiving fellowships.

Other cutbacks include the elimination of three major programmes: science education resources improvement (\$21.8 million in 1980), aimed at improving undergraduate instruction in two- and four-year colleges by upgrading instructional equipment and teacher competency; science education development and research (\$13.8 million) for developing new and more effective materials and methods of instruction in science and engineering; and science education communication (\$8.5 million).

Reaction to the proposed cuts has been sharp. Dr Alan Bromley, professor of physics at Yale University and president of the American Association for the Advancement of Science, warned that "to expect scientific and technological progress while abandoning efforts at improving science and engineering teaching in our schools is illogical and a disservice to the nation's interest". The association had previously announced that science education was to become a principal target of its own efforts over the coming year.

Dr Gerald Lieberman, dean of research at Stanford University, last week voiced the universities' anxieties: "In the past we have had to put little of our general funds to the support of graduate students in

Argentiniens face charges

Dr Jose Federico Westerkamp, formerly professor of physics at the University of Buenos Aires, was arrested on 27 February 1981 during a police raid on the human rights organization CELS (Centro de Estudios Legales y Sociales, Centre for Social and Legal Studies). He was held in custody for a week, then released. Police investigations into his human rights activities are, however, continuing.

Dr Westerkamp, who had been active on behalf of human rights in Argentina for several years, was a co-founder of CELS, together with two lawyers — Dr Emilio Fermin Mignione and Dr Augusto Conte Mac Donell. The group, which was set up early in 1979 and legally registered in March 1980, investigates human rights abuses, including the defence of political prisoners and the gathering of data on "disappeared" persons. Dr Westerkamp, whose son Gustavo has been a political prisoner since 1975, has been particularly active on behalf of imprisoned and "disappeared" scientists and academics, and has on numerous occasions spoken on their behalf at international scientific meetings.

According to the Argentinian newspaper, *La Razon*, CELS members face charges under article 224 for possessing plans and diagrams of military installations, carrying a possible penalty of two to eight years imprisonment.

science and engineering, but it is likely that we will now have to reevaluate our whole policy in this area". Dr John C. Crowley, of the Association of American Universities, claims that the cuts would wipe out "one of the two functions for which the foundation was established" and "calls into question the basic purpose of the National Science Foundation".

At the foundation itself, where it has been frequently claimed that science education has received less support than basic research, officials admitted that the cuts were "hard to square" with the report presented to President Carter last autumn.

Federally-sponsored science education programmes have some powerful opponents on the right. Some argue that, in principle, the federal government should play a minimal role in curriculum development, leaving it to state and local schools boards; others have criticized the foundation's education activities in the past, for example in developing biology and social science courses which, they claim, reinforce "relativist" values by portraying man and woman as part of nature, rather than unique creations. The cuts in the science education budget seem to have been orchestrated by groups such as the conservative Heritage Foundation and are likely to receive support from conservative members of Congress.

David Dickson

Swiss nuclear power

Sanctions bite

Zurich

There is growing concern in Switzerland about uranium supplies for the nuclear energy programme. Indigenous deposits are too small to reduce foreign dependence, and the efforts of the chief suppliers, Canada and the United States, to prevent nuclear proliferation have led to acts which the Swiss consider to verge on blackmail. Although the pro-nuclear lobby maintains that the long-term prospects are "not unfavourable" and that the present supply hitch is merely a temporary inconvenience, there are signs that the country's independence, self-determination and neutrality will be threatened if nuclear power becomes a major element in its energy mix — just as today they are threatened by an overwhelming dependence on imported oil.

The nuclear plan — yet to go through parliament — foresees a capacity of about 7,000 MWe in conventional nuclear reactors by the beginning of the next century. With an average reactor life of 30 years, total requirements of uranium will be around 40,000 tonnes. A laborious search for indigenous deposits over the past 20 years has turned up several occurrences of interest, none of them economically viable.

The highest grade mineralizations occur in the Vallorcine granite in the Vallée du Trient (Valais) with average uranium contents of 0.1 to 0.2 per cent. But extensive exploratory drilling and trial shafts revealed reserves of less than 2,000 tonnes. Larger reserves are suspected in the upper Rhine valley near Trun (Graubünden), but there the average grade is too small (0.02–0.1 per cent). An unpublished government report in 1977, after an initial exploration costing 6 million francs (£1 = SF 4.33), concluded that the deposits in the Alps can be regarded only as emergency reserves. Normally, Swiss reactors will be completely dependent on imported fuel. It is this dependence which is beginning to cause some alarm.

The present dispute with Canada illustrates the issues. Since 1958, Canada has been the main supplier of natural uranium under a bilateral treaty. In 1977, however, the Canadian government refused permission for the transfer of yellow cake to the fuel fabrication plants in the United States which the Swiss utilities had already paid for. Since then, almost 1,000 tonnes of uranium have been blocked, and the Swiss utilities have been forced to buy alternative supplies on the international market — at far higher prices.

The reason for the blockade is the reluctance of the Swiss government to give in to Canada's demand for inspection and control rights with respect to some Swiss exports. Since the explosion of an Indian nuclear device in 1974, Canada has insisted

that every effort must be made to prevent sensitive materials, goods and technologies of "essentially" Canadian origin from being passed on to third countries without Canadian permission.

The Swiss government is in an extremely awkward position. The Bundesrat is being pressured from all sides and is still vacillating. The utilities consider Canada as their main long-term supplier and want the Canadian demands to be met at once, with as little fuss as possible. In fact, they have already negotiated the required amendments to the treaty, with rather furtive government backing. But this document has been gathering dust since 1979 because of its political explosiveness. The opposition by industrialists and politicians to a treaty modified under duress is expected to be loud and strong. The former see it as a completely unwarranted interference in the free market and object to controls not governed by Swiss laws. The latter see the loss of sovereignty and capitulation to outside pressure as matters of principle. Jurists have even suggested that Canada's failure to supply is a breach of the treaty which should be taken to the international courts — an interpretation which the Bundesrat has vehemently denied. And anti-nuclear members insist on asking awkward questions in parliament.

The general public, which is already rather sceptical of the future of nuclear energy, is confused. Industry and government experts maintain that the situation is under control and that it is only an initial

difficulty which, once mastered, will be followed by long-term guarantees of fuel supply. International cooperation and the direct financial participation of the utilities in uranium exploration will, they say, ensure that future requirements will be met. But the signs are hardly encouraging.

Two tendencies suggest that uranium mining and export will become increasingly controversial issues in the internal politics of the producing countries — the inevitable exploitation of lower grade ores, resulting in greatly increased environmental impact, and the increasing concern about the spread of nuclear weapons. The rapid development of nuclear power elsewhere will keep the world uranium market under pressure — and minor users like Switzerland at a constant disadvantage. If Switzerland's uranium supply today is being adversely affected by the explosion of a bomb nine years ago on another continent, what next?

Geoff Milnes

Indian community health

Cure for all ills

Lucknow

The Indian government's target of "health for all" by the year 2000 will not be met unless present policies are radically changed, according to a study group set up by the Indian Council of Medical Research (ICMR) and the Indian Council of Social Science Research (ICSSR).

Their report stresses that health should be recognized as a function of the overall development of society. An "alternative health care model" is presented which includes programmes for national economic growth with the objective of doubling national income per capita by the year 2000, full-scale employment at reasonable wages, improvement in the status of women, adult education, elementary education for all children between 6 and 14, welfare of backward classes, rural electrification, housing and organization of the poor. The report also suggests setting up a National Population Commission with the objective of stabilizing the total population at 1,200 million by the year 2050.

The "alternative health care model" encompasses promotive, preventive and curative aspects and combines "the valuable elements in Indian culture and tradition with the best elements of the Western system".

The report suggests that a Medical and Health Commission should be set up to look after the quantitative and qualitative aspects of health education. It recommends that drug production should be correlated with disease patterns with a bias towards the economic production of basic and essential drugs. On the research side, the important areas are delineated as parasitic diseases, problems of environment, sanitation and indigenous medicine.

Zaka Imam

More interferon cloned

The interferon genes of α and β interferon, produced respectively by lymphocytes and fibroblasts, have been cloned successfully in *Escherichia coli* at the University of Warwick, according to an announcement this week from the university. The project, led by Professor D.C. Burke, has been under way for the past three years. The Wellcome Foundation, one of several sources of finance, is also collaborating technically.

Although the two interferon genes have been successfully cloned in a bacterial plasmid, the genes have not yet been incorporated at plasmid sites where they are expressed, so as to produce the protein interferons.

Although at least three other groups (Biogen, Genentech and G.D. Searle) have successfully cloned interferons, doubts about the patentability of the techniques imply that being fourth is not necessarily to be robbed of reward.

For cloners of interferon, however, the most obvious prize for the immediate future is the cloning of gamma-interferon, otherwise known as immune interferon, which may have more interesting physiological properties than the two now in clinical trials.

Oil tanker hazards

Inaction at sea

Brussels

The recently published Liberian report on the *Amoco Cadiz* disaster draws attention to the belief widely held in shipping circles that 80 per cent of accidents at sea are due to human error. In March 1978 the tanker went aground off the coast of Brittany spilling 230,000 tonnes of oil. And now the captain is to get his licence back despite his failure to send out a distress signal at the earliest opportunity.



Amoco Cadiz — what happens next time?

The incident prompted the European Commission to spend two and a half years in preparing a set of proposals on how such accidents could be prevented and cleaned up in the future. The proposals were published last June, and despite pressures from the French government they run the risk of never being accepted.

The most controversial issue concerns a requirement that member states should identify sub-standard ships using their own ports and order deficiencies to be rectified. Now that Greece has joined the European Community, 30 per cent of the world's shipping falls under Community jurisdiction. A recent debate in the European Parliament on the idea that the port state should check the standards of shipping ended in uproar, with the Greek representatives defending themselves against accusations that most defective ships were Greek-owned.

Conventions on shipping safety have already been established by the Inter-governmental Maritime Consultative Organization (IMCO), but signatory countries have been slow to enforce them. The Commission's directive would make the IMCO conventions mandatory.

Dissatisfaction at the slow progress of both IMCO and European Commission proposals led France to hold a conference on shipping safety and pollution in

December last year. The Paris conference, which was attended by Spain, Portugal and Nordic countries as well as the European Community, set up a committee, which has been meeting in the Hague, with the declared aim of drawing up rules on the obligations of port states which would be acceptable to all European countries.

While these deliberations go on, the European Commission has made still more proposals, intended to stand more chance of being adopted quickly. One directive, already adopted, calls for all relevant data on tankers and their movements to be

stored on a computer. Another requires details of the contingency plans of all member states to be computerized, and the Commission wants to run training schemes and equipment testing programmes. Many such precautions are already in operation in the United States, causing operators with sub-standard ships to concentrate on European trade.

Whatever the various European Community bodies decide, nothing much is likely to happen for some time. The United Kingdom and Denmark still refuse to countenance giving the Community competence in this field.

In the meantime, France is still pressing for the £670 million estimated to be the cost of damages caused by the *Amoco Cadiz*.

Jasper Becker

US animal research

Stricter safeguards

Washington

Although the United States has so far avoided violent demonstrations against the use of animals in research such as those in Canada and the United Kingdom, public pressure for stricter safeguards is mounting. Three separate bills have been introduced in the new session of Congress. One, first introduced by Representative

Frederick Richmond in 1979, would permit the withdrawal of public funds from any research project in which laboratory animals were judged to be used unnecessarily.

The research community, and officials at the National Institutes of Health (NIH), are dismayed at the prospect of new regulations. At a meeting last month convened by NIH, scientists and animal welfare groups discussed the feasibility of the wider use of *in vitro* testing methods and mathematical modelling as alternatives to the use of live animals. Dr William Raub, associate director at NIH, said that he would approach the Office of Science and Technology Policy (if it survives under the new Administration) to suggest a regular forum for the discussion of alternatives to animals.

This proposal should at least satisfy Congressman George Brown, previously chairman of the House of Representatives Science and Research Committee, who came out in favour of more open discussion between scientists and their critics after a flood of correspondence supporting the Richmond bill. One of Mr Brown's objectives is to avoid trouble such as that in Canada, where radical anti-vivisectionists last year broke into National Research Council laboratories, setting free laboratory animals.

In the United States, protests against the widespread use of laboratory animals have taken the form of lobbying, led by groups such as United Action for Animals, based in New York.

One result so far stems from the campaign last year by more than 400 animal welfare groups against the Draize irritancy test for cosmetics, in which liquids are dropped on rabbits' eyes to test for inflammation. Revlon, one of the largest cosmetic manufacturers, was singled out for criticism and last November gave a \$750,000 grant to Rockefeller University to investigate cell culture alternatives.

The passage of new legislation will not, however, be straightforward. Present legislation goes back to the 1960s and is concerned with the care of laboratory animals, not with the tests to which they are subjected. The Scientists' Center for Animal Welfare is pressing for the wider adoption of a procedure already introduced by the Veterans Administration in which research proposals are vetted in advance to ensure that appropriate procedures are followed in the use of laboratory animals.

Mr Richmond's bill, the front runner in Congress, would establish a centre for the development of methods of research and testing not involving the use of live animals. Supporters of the bill say that such a centre could operate in much the same way as the National Toxicological Program, introduced three years ago, which coordinates research on toxic substances in the major federal agencies. The centre would have power to direct between

30 and 50 per cent of NIH funds for animal-based research into the development of alternative testing methods.

Critics, such as the National Society for Medical Research, argue that the bill would give federal regulators "potentially catastrophic" powers and that scientists use alternative methods voluntarily when they are available. Feelings on both sides run high, however. NIH officials, previously cool to the demands of the animal welfare groups, seem now to be hoping that if they can bend with the wind, legislation will be avoided.

David Dickson

UK animal research

Legislation unlikely

British legislation on laboratory animals is hanging fire until the draft of a Council of Europe convention is agreed, supposedly in May. The government has said that new legislation will not be based on the amended version of the private member's bill first introduced into the House of Lords by Lord Halsbury last session and reintroduced this session. Members of the Lords committee that sat to amend that Halsbury bill now fear that further delays, due to tardiness in the Council of Europe or in drafting a new bill, could jeopardize agreement reached by opposing sides during the committee's deliberations.

The government hopes to avoid passing legislation clashing with the convention's aims. British legislation, however, is bound to be more stringent than the lowest common denominator between the Council's member states laid down in the convention. The latest available draft contains nothing that contradicts the Halsbury bill or that is likely to give the government's bill drafters pause.

Whether further amendments will cause problems remains to be seen. But the government's objections to the Halsbury bill are clearly on other grounds. During the second reading debate before Christmas, Lord Belstead for the government said that the bill leaves too much to the discretion of the Home Secretary and invests too many powers in an advisory committee that would be set up to review regularly ethical matters and the replacement and use of laboratory animals.

Lord Halsbury is far from content. He suspects that the main objections to his bill have come from Home Office inspectors who fear that an advisory committee, along the lines laid down in the bill, would disrupt their well-tried methods of working. Despite government assurances that the original timetable still stands, Halsbury fears that delays in the Council of Europe could push it back. He plans to keep the issue alive by pushing his amended bill through the committee and report stages in the Lords in May even though it is now clear that it will never get a reading in the Commons.

Judy Redfearn

Commonwealth meeting

Food for reserves

Dacca

Moves to bolster the dangerously low level of world food stocks were the main outcome of a meeting of Commonwealth ministers for agricultural and rural development held in Dacca, Bangladesh, on 11–13 February. The ministers agreed that by mid-1981 negotiations should be completed to create a new International Grains Arrangement aimed at providing some of the 500,000 tons of cereals needed annually to replenish the International Emergency Food Reserve.

This was the first full-scale Commonwealth ministerial meeting of its kind, and was attended by 21 of the 44 member countries. The Commonwealth is made up of 20 nations showing a food deficit, and 21 with a surplus in food production. The remaining 3 nations, including host-nation Bangladesh, are on the verge of becoming self-sufficient.

The ministers urged quick action from international funding agencies, especially the World Bank and the International Fund for Agricultural Development, which are concentrating on aid to the poorest countries. Extra resources were called for to enable the Commonwealth Fund for Technical Cooperation to give more aid to developing Commonwealth countries.

M. Kabir

Romanian agriculture

Research expansion

President Nicolae Ceausescu of Romania called last month for greater research effort in agricultural sciences. He was addressing the 11,000 delegates at the Second Congress of Workers in Agriculture, including experts on agriculture as well as workers from cooperatives, collectives and the small residual private agricultural sector.

Although, according to Angelo Miculescu, the Minister of Agriculture, agricultural production went up by 26.4 per cent during the 1976–80 Five Year

Romanian workers — due for some changes

Plan, President Ceausescu had previously pointed out, at a working meeting on agriculture in January, that targets during the past five years had not been fulfilled, and that "resolute" measures would still be needed. Such measures, he has now told the congress, must include expansion of research on plant and animal genetics, soil quality and, for the economists, the logistics of a switch from the traditional farming structure of the countryside, to the development of agro-industrial complexes. The latter scheme, by which food-processing plants are located in the countryside, not only eliminates the need of transporting perishable foodstuffs to the cities for processing, but provides additional jobs to reduce the drift to the cities.

The new research priorities will be met from the present budget, apparently by switching funds from other sectors.

In all these changes, said President Ceausescu, the Academy of Agricultural Sciences in Bucharest must have an increased involvement both in planning and coordination. The Academy, and, indeed, the whole educational system, he continued, must train more agricultural specialists.

Higher education in the agricultural sciences, however, is not to be open to all. President Ceausescu says that only workers or former workers in agriculture may go to university to study agriculture. The practice by which a young city-dweller can go to an agricultural institute simply to qualify for a clerical job nominally connected with agriculture is to stop.

Similarly, agronomy centres where specialists are trained are to be reorganized. At present, future experts are trained at "picturesque" centres, whose sites are frequently chosen for reasons other than agricultural suitability. At one such centre near Bucharest, for example, experts are trained in handling tractors in hilly areas, although the centre itself is remarkably flat. The reason is simple, said the president. "Comrades from the Ministry of Agriculture" want the centre there, because it is a convenient place for them to go and make speeches.

Vera Rich



CORRESPONDENCE

Genesis of genetics

SIR — I find the current interest among biologists in the findings of Górczynski and Steele^{1,2} somewhat surprising, since a report of Lamarckian inheritance was published about 3,000 years ago. It is written that:

"And Jacob took him rods of green poplar, and of the hazel and chestnut tree; and pilled white strakes in them, and made the white appear which was in the rods. And he set the rods which he had pilled before the flocks in the gutters in the watering troughs when the flocks came to drink, that they should conceive when they came to drink. And the flocks conceived before the rods, and brought forth cattle ringstraked, speckled, and spotted."³

Were these findings ever reproduced? If so, would it not be more appropriate to speak of Jacobinic genetics?

KIRSTEN FISCHER LINDAHL

Basel Institute for Immunology,
Basel, Switzerland

1. Górczynski, R.M. & Steele, E.J. *Proc. natn. Acad. Sci. U.S.A.* 77, 2871-2875 (1980).
2. Górczynski, R.M. & Steele, E.J. *Nature* 289, 678-681 (1981).
3. *Genesis* Ch. 30, v. 37-39.

Ethics of genes

SIR — It was with considerable surprise and no little confusion that I read Richard Dawkins' letter on genetic determination (*Nature* 12 February, p.528). In his commendable desire to dissociate himself from the National Front, he has left me totally perplexed about his actual views of the relation between genotype and phenotype. Near the end of his letter, he associates himself with the views of S.J. Gould, that the genetic basis of IQ is "trivially true, uninteresting, and unimportant". Yet earlier in the same letter, he says that genetics is sort of relevant since we may need to "fight all the harder" against genetic tendencies. But in his book *The Selfish Gene*, Dr Dawkins wrote that we are "robot vehicles blindly programmed to preserve the selfish molecules known as genes" (preface) and that these genes "swarm in huge colonies, safe inside gigantic lumbering robots, sealed off from the outside world . . . manipulating it by remote control. They are in you and me; they control us body and mind" (p.21).

It really is very vexing. Just as I had learned to accept myself as a genetic robot and, indeed, felt relieved that I was not responsible for my moral imperfections, Dr Dawkins tells me that, after all, I must try hard to be good and that I am not as manipulated as I thought. This is a problem I keep having in my attempt to understand human nature. Professor Wilson, in his book on sociobiology, assured me that neurobiology was going to provide me with "a genetically accurate and hence completely fair code of ethics" (p.575). I was

euphoric at the prospect that my moral dilemmas at last had a real prospect of resolution, when suddenly my hopes were dashed by an article in which Professor Wilson warned me against the naturalistic fallacy (*New York Times* 12 October, 1975). You can imagine my perplexity. I do wish I knew what to believe.

Perhaps I am just asking for that foolish consistency which Emerson tells us is the hobgoblin of small minds. But I see that Dr Dawkins himself is uncertain. I can only echo the question he asks in his letter. "Where on earth did the myth of the inevitability of genetic effects come from? Is it just a layman's fallacy, or are there influential professional biologists putting it about?"

ISADORE NABI

Museum of Comparative Zoology,
Harvard University, Cambridge,
Massachusetts, USA

Cancer causation

SIR — The Epstein-Swartz article which appeared in *Nature*¹ (and was essentially identical to the manuscript these authors presented at the autumn meetings of the American Public Health Association in Detroit) will elicit a variety of critical responses from those in the scientific community who are familiar with the specific aetiological links between cancer and the environment.

But here I would like to point out one major fallacy in the Epstein and Swartz piece — their premise (that lifestyle factors have been overstated in discussions of the aetiology of cancer) is based largely on the assumption that occupationally-induced cancers are more prevalent than previously thought. Epstein and Swartz criticize Dr Peto for dismissing "recent estimates of the importance of occupational carcinogens in a report by the US Public Health Service as exaggerated, unsound and unreasonable". This alleged "report" concluded that "as much as 20% or more" of cancers in the near term and future may reflect past exposure to specific carcinogens in the workplace.

Epstein and Swartz attempt to build up the credibility of this estimate by stating that the report was "prepared by nine named and internationally recognized experts in cancer epidemiology, statistics and carcinogens from three federal research agencies" and go on to add that "there is no basis whatsoever for recent unsubstantiated allegations by Peto and others that all or most of the authors of the government report have disowned or rejected it or its conclusions".

Epstein and Swartz are in error on two points in their discussion of occupational cancer and the so-called "estimates paper". First, nobody except those two authors and perhaps some regulators who wished to bolster their requests for more legal control over

occupational chemicals has ever cited the "20%" estimate in a serious, scientific context. Second, according to a detailed article by Ruth B. Schwartz in our publication *ACSH News and Views*² "7 of the 10 scientists responsible for these statistics believe they may be incorrect. Only one scientist defends them as reasonable". Two could not be reached for their opinions. Ms Schwartz's conclusions were based on her in-depth discussions with the various "contributors".

Perhaps the most succinct evaluation of the "estimates paper" on which Epstein and Swartz rely so heavily was made by Sir Richard Doll: "I regard it (the "estimates paper") as scientific nonsense".

ELIZABETH M. WHELAN

American Council on Science and Health,
New York, USA

1. Epstein, S.S. & Swartz, J.B. *Nature* 289, 127-130 (1981).
2. Schwartz, R.B. *ACSH News and Views* Nov/Dec (1980).

Badger debate

SIR — Dr Flowerdew's recent letter (*Nature* 26 February, p.742) makes me wonder what purpose the Mammal Society intended to serve by engaging in what can indeed now be called "the badger controversy". Dr Flowerdew says that the society spoke out because my report provided a "one-sided" interpretation of the evidence ("biased" in Dr Harris's letter of 11 December; *Nature*, p.532), and because my "categorical" conclusions and recommendations did not take sufficient account of "the complexity of the problem". But in Dr Harris's letter it was also said that the society was intervening because it was necessary to correct certain "factually misleading statements" that I had made. Yet neither letter adduces any new "fact". Those that they use are the ones which I had assembled. Nor did either correspondent spell out any "anomalies" in the story other than those which I had considered. Dr Flowerdew now says that there is no unequivocal evidence that the gassing policy will produce a long-term solution, by which I presume he means stamp out the disease. But no one has ever claimed that it would. Above all, the Mammal Society, on whose behalf Drs Flowerdew and Harris have written, does not propose any alternative policy which the Government could pursue in order to lower the level of the reservoir of the bovine tubercle bacillus in the affected areas of the South West — the existence of which, as Dr Plowright pointed out (*Nature* 1/8 January, p.8), they first denied, but which in Dr Flowerdew's present letter he, or the Mammal Society, now seems to accept.

I intended no "attack on the Mammal Society and some of its members" — Dr Flowerdew's words — when I asked on what scientific authority the society based its presumed criticisms. The authorities on whom

I myself relied were first, the many scientists and veterinary surgeons from the Ministry of Agriculture, Fisheries and Food whom I cross-examined during my enquiry; second, the three internationally-known students of animal disease whose help I acknowledged in my report — Sir William Henderson, FRCVS, FRP, Dr L. Goodwin, FRCP, FRP, and Dr W. Plowright, DVSc, FRCVS; and third, Professor T. McKeown, FRCP, well known for his studies of trends in tuberculosis and other diseases; the authorities in the Medical Research Council and Department of Health and Social Services who are concerned with tuberculosis; and Dr B.R. Cook, until recently with the New Zealand Ministry of Agriculture.

The officers of the Mammal Society may believe that their society is exempt from the usual convention that no scientific society has the licence to make corporate statements of a scientific nature on behalf of all its members. But if it is exempt, one is entitled to ask — indeed, one is in duty bound to ask, since we are talking about a matter of animal and public health that is of international concern — who were the authorities on whom the Mammal Society leant, and whose counsel was denied me. Dr Flowerdew bluntly declares that the officers of the society were responsible for the statements that have appeared over his and Dr Harris's name. But apart from Dr Frazer, not one of the officers listed by the society is shown as having either a medical or veterinary qualification, and while some of its members, in particular Dr Neal, are well-known badger naturalists, none has written — unless in some unquoted journal — either on TB in cattle or on disease in badgers. Indeed, apart from Dr Neal's writings, I can find only one paper on badgers by any officer of the society, and that is a note by J.F.D. Frazer on badgers in Kent. The society's publications appear in the *Mammal Review* and in the form of "notes" in the *Journal of Zoology*, published by the Zoological Society of London.

Four of some 120 papers that have appeared in the lifetime of the *Review* concern badgers, and only seven deal with pathological matters (mainly ectoparasites — none deals with TB or with diseases in badgers); while only 15 of 311 "notes" concern badgers, and not one TB. In the circumstances, I find it more than a little strange that Dr Flowerdew suggests, presumably on behalf of his society, that the Ministry of Agriculture, Fisheries and Food should establish a Scientific Advisory Group "containing a small number of independent scientists such as wild-life epidemiologists and statisticians" — he offers no names — to "analyse the data in detail" for the ministry's Consultative Panel on Badgers and Tuberculosis, with the unfortunate imputation that in the view of the Mammal Society the ministry's scientific and veterinary officers are incompetent, and that their analyses of the data they collect cannot be relied upon.

LORD ZUCKERMAN

University of East Anglia,
Norwich, UK

SIR — Why has the Mammal Society (*Nature* 26 February, p.742) not attempted a reasoned reply to the criticisms contained in the letters by Dr Plowright (*Nature* 1/8 January, p.8) and myself (*Nature* 22 January, p.218), and by Zuckerman in his further article (*Nature* 19 February, p.628)?

If it is now accepted that the use of gassing

by the ministry has substantially reduced herd breakdown in affected areas, its continuation is clearly indicated. There does not at present appear to be any other way of tackling this source of infection. Reduction of badger densities in affected areas not only reduces the chance of infecting cows because there are fewer badgers around, but offers the best hope of greatly reducing and possibly, as more is learnt, substantially eliminating tuberculosis from the badgers themselves. In this respect badgers are fortunate in that there is now no major external source of infection and the TB monitoring of dairy herds gives early warning of the presence of infected setts in a locality.

In their earlier letter the Mammal Society, while suggesting that "other more subtle factors may be significant", did not dispute that density might be of importance in the spread and maintenance of TB in the badgers. They have now gone back on this, stating that "there is no unequivocal evidence that TB in badgers is density-dependent". This statement is based on the maps of badger density and herd breakdowns included in the Zuckerman report (p.68,70) which, they say, show that "TB is not only prevalent in areas of high density, but also in areas such as South Dartmoor where the badger population is very low". Had they referred to the subsidiary reliability diagram relating to the badger density map they would have found that the cluster of infections in South Dartmoor falls in an area classified on the reliability diagram as "mainly guesswork". They have completely ignored Zuckerman's discussion in his article of the statistical laws governing the propagation of infectious diseases, and his deduction that the establishment and maintenance of TB among badgers in an area is likely to be strongly density dependent, from which it follows that, as a matter of long-term policy, badger population densities should not be allowed to become excessive, as they certainly appear to be in parts of the South West at present.

The reactions to our other criticisms are similarly irresponsible and misleading. They dismiss Dr Plowright's letter with the comment that it "presented no new information nor did he answer any of the points we raised". The main point of my letter was to cast doubt on the Mammal Society's suggestion that the apparent success of the gassing campaign might be due to the decline of TB throughout the country. They appear now to have partially, but not wholly, conceded this point. They make much of the large decline from 5.5 to 3.2 per cent of herd breakdowns in Cornwall from 1975 to 1976 "highlighted" by my graph, and claim that it could not have been wholly produced by the gassing campaign. That there was a marked rise all over the country in 1975, followed by a corresponding decline in 1976, was clearly shown by my graph, and was pointed out in my letter, but this does not preclude that a large part of the Cornish decline was due to the removal of infected badger groups, some of which may have been done by farmers without the cooperation of the ministry. In any case, if the Mammal Society wished to comment on the irregularities of the Cornish data they might better have directed attention to the disturbing rise of infection between January 1977 and September 1979 which, however, was matched by a substantial fall between October 1979 and March 1980.

There are two further glaring distortions of the evidence in their current letter. The first is

my "contention", for which I gave reasons, that the 15 per cent of Cornish herd breakdowns attributed to badgers in the report was clearly much too low. That this contention was indeed correct was substantiated by the further information that Lord Zuckerman gave in his article. The Mammal Society accept this, but by lumping together the figures for 1974–75 and 1976–78 they conceal the progressive reduction in the proportion of "unknowns" and were able to claim that "since 1974 badgers were still only believed responsible for a minority of breakdowns".

A similar concealment of the full evidence occurs in their comment on the origins of infection from specific badger setts. Zuckerman reported that he had been informed that for the whole of Cornwall 72 per cent of the 51 outbreaks definitely attributed to badgers were within 2 miles of infected setts and 33 per cent were within half a mile, and that for the West Penwith area the corresponding percentages were 97 and 47. All that they state is that for the whole of Cornwall 28 per cent (100 minus 72) were attributed to setts more than 2 miles away and that badgers rarely travel this distance in areas of high density. A glance at the herd breakdowns map (p.20) of the report shows that the West Penwith area, which covers the western tip of Cornwall, contains the major pocket of infection and was clearly the source of the majority of the herd breakdowns.

Finally, in connection with their long and almost unintelligible argument on the effects of the moratorium on gassing, may I make it clear that neither Zuckerman nor I ever argued that "TB in badgers and cattle immediately increased as a result of the moratorium"; we merely reported the latest available data. Obviously there are time lags. The full effects of the moratorium cannot be properly assessed until data for a further year are available. Also in this connection, what is one to make of the statement: "This . . . totally ignores the likelihood of natural cyclic trends in TB prevalence, as discussed in our original letter and the reply by Dr Yates"? There is no mention of cyclic trends either in their letter or in my reply!

FRANK YATES

Rothamsted Experimental Station,
Harpenden, UK

SIR — You have recently drawn attention to the controversy over the badger-gassing issue.

The Zuckerman report shows a rise and fall of tuberculosis in the South West and also in the rest of England from 1974 to 1980. Some factor, common to the South West and the rest of England, must therefore be sought. I believe that factor is one which the Zuckerman report carefully avoids — the resumption of health-certification of the Irish cattle imports in June 1976 after nearly two years of importing TB-infected cattle into the United Kingdom during the Irish veterinary inspectors' strike.

The reluctance of the Animal Health Authority to suspend Irish cattle imports at that time contrasts sharply with their present enthusiasm to destroy badgers on scant evidence.

PETER H. ROBERTS

Compassion in World Farming,
Petersfield, UK

NEWS AND VIEWS

Propagating rifts in seafloor spreading patterns

from Tanya Atwater

SPECIALISTS from diverse geological fields recently gathered together to discuss mutual studies in the Cascadia region of western North America. One of the highlights of the meeting was the showing of a computer simulation of propagating rifts in the Juan de Fuca region by Richard Hey of the Hawaii Institute of Geophysics. Ever since Vine and Wilson¹ applied their seafloor spreading description to the well mapped magnetic anomalies of this region, enthusiasts have been cutting and pasting coloured maps, trying to explain the oblique fractures and other complexities in the pattern. For several years Hey and his colleagues have been developing and refining the notion of propagating rifts^{2,3} to explain data collected along the relatively remote Galápagos spreading centre. When they applied the concept to the Juan de Fuca anomalies, the many complexities became easily understood and predictable.

Probably the most convincing evidence for seafloor spreading, continental drift and plate tectonics lies in the configuration of linear magnetic anomalies from the sea floor. Vine and Matthews⁴ and Vine⁵ showed these anomalies to be indicators of symmetrical additions of new material along mid-ocean ridges and Wilson⁶ and Vine and Wilson¹ showed that these anomalies are offset along transform faults and their fossil traces, fracture zones. The now classic pattern of anomalies and offsets is that shown in Fig. a. This pattern is constructed assuming that the transform

offset was inherited from the original shape of the spreading boundary and that it constitutes impenetrable end points for the two ridge segments which abut it.

The simple pattern of Fig. a is a useful model both for explaining the gross configurations of seafloor features and for determining world plate motions. For fine details, on the other hand, the pattern is an oversimplification. Anomaly profiles commonly show occasional local sections of extra or missing space compared with predicted symmetrical patterns. These perturbations have generally been attributed to small 'ridge jumps' at which time the centre of spreading suddenly shifted sideways to a new location, breaking a section off from one plate and adding it to the other. In interpretations of sparse magnetic profiles, the detected spreading centre jumps have generally been assumed to be instantaneous breaks across new ground. More dense surveys have shown this not to be the case, however. Rather, the new ridges appear to tear through the crust at rates comparable to spreading rates, 'propagating' forward from their tips and gradually overlapping and extinguishing the previously existing spreading centres.

The magnetic pattern created by such a propagating rift is most easily visualized in its stepwise form, as shown in Fig. b after Hey². At regular intervals, the lower right-hand rift propagates past the transform fault, taking over the spreading for a new length of ridge and establishing a new

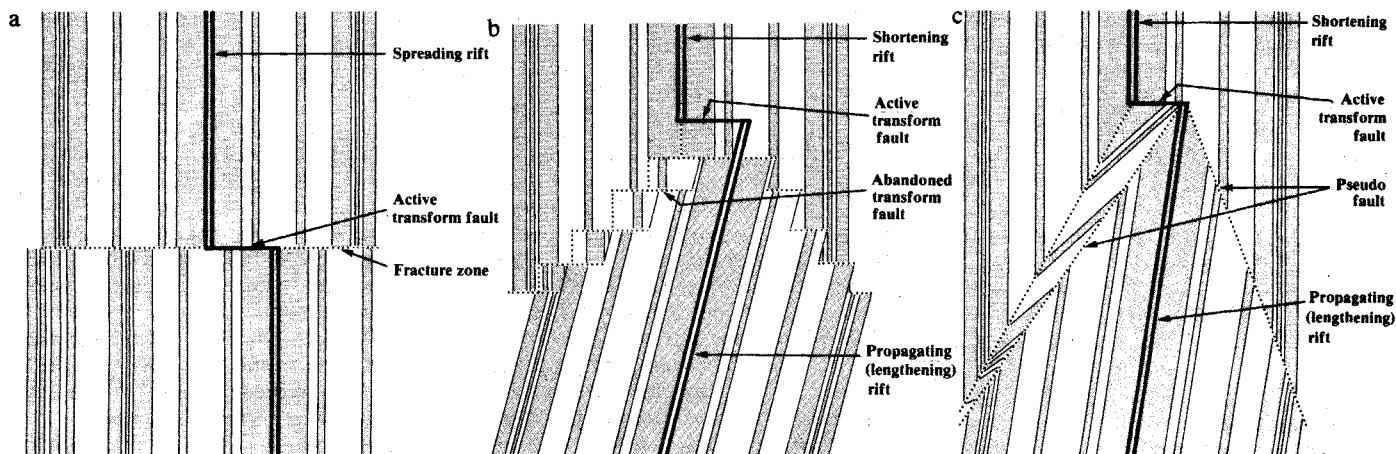
transform fault at its tip.

The third figure, after Hey *et al.*³, shows the anomaly pattern formed by a continuous, slow propagation of the lower rift and continuous shifting of the connecting transform zone. The oblique V-shaped traces left in the anomaly pattern appear at first sight to be fault lines along which the anomalies were broken and offset after their formation. Hey² notes that, in fact, no fractures ever occurred in these directions. They are inherited patterns which he calls 'pseudofaults'. V-shaped patterns like these are very clear in the Juan de Fuca magnetic surveys⁷, in the Surveyor surveys further offshore⁸ and in the Galápagos surveys^{3,9}. In the Atlantic, magnetic anomalies are much less clear because of slow spreading rates, but V-shaped patterns have long been noted in the topography near the Reykjanes ridge¹⁰ and at mid latitudes¹¹.

Another phenomenon clearly related to the propagators is the occurrence of ferro basalts and related high-amplitude magnetic anomalies¹². These unusual basalts erupt primarily near the growing rift tips and may help us to understand the special environment in which these rifts occur.

The reason why one rift propagates to supercede another is not clear. It is common for the new rift to have a slightly different trend than the old, suggesting this is the mechanism by which spreading centres adjust to changes in stress directions. However, the propagation of

Marine magnetic anomaly patterns generated by seafloor spreading at: (a) A stable spreading-transform fault system¹; (b) An oblique rift propagating northwards in discrete steps²; and (c) An oblique rift propagating steadily northwards³. Oblique 'fractures' mapped at the Juan de Fuca and other spreading centres may actually be pseudofaults as shown in (c).



rifts can be seen to accomplish other interesting results, as well. In the Galápagos area, for instance, the rift has 'jumped' south several times (via propagating rifts), thereby maintaining a location near the Galápagos hot spot. Another suggestive relationship is that the propagators commonly migrate outwards from the proximity of hot spots³.

The examples of propagators thus far documented occur at transform faults of small offset. However, Hey *et al.*³ suggest that all rifts begin as propagating cracks. The present East Pacific Rise in the South Pacific seems to have slowly grown, gradually replacing one that lies, now extinct, to the east¹³. Similarly, the original rifting of the Atlantic may have been progressive as suggested by the non-isochronous edge of the magnetic quiet zone¹⁴. A more recent version of continental rifting can be seen in the Gulf of Aden and Gulf of Tadjourah where the rift appears to have propagated westwards

Tanya Atwater is Professor of Marine Geophysics at the University of California, Santa Barbara.

into the African continent at a rate of about 30 mm per year¹⁵.

The propagating rift idea is really only a small modification to our understanding of seafloor spreading. Yet, it is an extremely satisfying one since it explains so many diverse 'imperfections' in the record in such a simple way, and it gives us a much better insight into the processes by which rifts form and readjust to changing conditions. □

1. Vine, F.J. & Wilson, J.T. *Science* **150**, 485 (1965).
2. Hey, R. *Earth Planet. Sci. Lett.* **37**, 321 (1977).
3. Hey, R., Duennebie, F.K. & Morgan, W.J. *J. geophys. Res.* **85**, 3647 (1980).
4. Vine, F.J. & Matthews, D.H. *Nature* **199**, 947 (1963).
5. Vine, F.J. *Science* **154**, 1405 (1966).
6. Wilson, J.T. *Science* **150**, 482 (1965).
7. See, Vine, F. for example, *History of the Earth's Crust* p.73 (1968).
8. Shih, J. & Molnar, P. *J. geophys. Res.* **80**, 4815 (1975).
9. Hey, R. & Vogt, P.R. *Tectonophysics* **37**, 41 (1977).
10. Vogt, P.R. *Earth Planet. Sci. Lett.* **13**, 153 (1971); Johnson, G.L. & Vogt, P.R. *Geol. Soc. Am. Bull.* **84**, 3443 (1973).
11. Rona, P.A. *Earth Planet. Sci. Lett.* **30**, 109 (1976).
12. Vogt, P.R. & de Boer, J. *Earth Planet. Sci. Lett.* **33**, 145 (1976).
13. Herron, E.M. *Geol. Soc. Am. Bull.* **83**, 1671 (1972).
14. Hayes, D.W. & Robinowitz, P.D. *EOS Trans. AGU* **56**, 450 (1975).
15. Courtillot, V., Galdeano, A. & LeMouél, J.L. *Earth Planet. Sci. Lett.* **47**, 144 (1980).

The detection of cellular transforming genes

from Peter W.J. Rigby

THE demonstration that the transforming genes of retroviruses are derived from cellular genes, designated proto-oncogenes¹⁻³, has considerable implications. If certain cellular genes are oncogenic when placed under the influence of viral control elements, perhaps other genes, or indeed the same ones, mediate chemical carcinogenesis by being expressed at high levels as the result of carcinogen-induced mutations. Recent work from the groups of Robert Weinberg (MIT) and Geoffrey Cooper (Harvard University) has shown that genes involved in chemical carcinogenesis can be detected by DNA-mediated gene transfer experiments.

Weinberg's group has been able to induce transformation by transfecting into non-transformed mouse fibroblast DNA from lines of mouse fibroblasts transformed *in vitro* by the chemical carcinogen 3-methylcholanthrene (MC)⁴. Efficient transfection occurs with only one such line, NIH3T3, the cells of which are presumed, for some reason, to be particularly efficient at taking up and expressing exogenous DNA. DNAs from fifteen transformed lines were tested and successful transfer of the transformed phenotype was obtained with DNA from five lines of MC-

transformed cells. High-molecular-weight DNA (~30 kilobases in length) was used and the efficiency observed was roughly 0.1 focus forming unit per µg of DNA (FFU µg⁻¹). Cell lines derived from the foci were subsequently shown to exhibit another parameter of transformation — the ability to grow in soft agar. Infection of the transformed cells with murine leukaemia virus failed to rescue a transforming virus, suggesting that transformation was not due to the activation of an endogenous retrovirus. The transformed phenotype could be passaged through a second round of transfection. Note that the DNA of the majority of cell lines tested could not transfer the transformed phenotype and, more important, neither could the DNA of some MC-transformed lines.

In a second report⁵, restriction endonuclease digestion of MC-transformed cell DNA was used to assess the number of transforming genes. A gene will be inactive in the transfection assay if an enzyme cuts within it whereas enzymes that cut outside the gene will have no effect. In each of four cases tested, *EcoRI* and *HindIII* abolished transformation whereas digestion with *BamI*, *XhoI* and *SalI* had no effect. It thus appears that the same gene is involved in transformation in each of the four cell lines. The transforming genes of other cell types, a rat neuroblastoma and a line of cells transformed by transfection with normal mouse DNA, exhibited

distinct patterns of restriction enzyme sensitivity. The lines showing the same pattern were all derived by *in vitro* transformation of mouse fibroblasts with MC. The DNA of many MC-transformed cells was not active in the transfection assay, so it seems likely that only a subset of MC-responsive sequences was being examined.

In this issue of *Nature* (p.261), Shih and his colleagues extend their observations by showing that the phenomenon is not confined to genes involved in the transformation of mouse fibroblasts. Cellular transforming genes can cross cell-type and species barriers — DNA from three out of four rat neuroblastomas induced by ethylnitrosourea transformed NIH3T3 cells as did DNA from a MC-induced mouse glioma. In each case the primary transfectants were tumorigenic and a second cycle of transfection could be demonstrated. Similar results were obtained with a variety of carcinomas, including the Lewis lung carcinoma of the mouse, a 7,12-dimethylbenzanthracene-induced mouse bladder epithelial carcinoma, a benzo-a-pyrene-induced rabbit bladder epithelial carcinoma and a human bladder epithelial carcinoma. The phenomenon was again not general, the DNAs of several neuroblastomas and carcinomas being inactive, and rescue experiments argued against the involvement of retroviruses. This paper provides direct evidence that mouse cells transformed by human carcinoma DNA have acquired human sequences. In DNA-mediated gene transfer experiments, cells generally seem to take up not only the gene that is being selected for (in this case the carcinoma transforming gene) but also any other DNA that happens to be in the calcium phosphate precipitate. Human DNA contains a repeated sequence family, known as the *Alu* family, that is not detectable in the mouse genome. Using a cloned *Alu* fragment as hybridization probe it was shown that the transformed mouse cells had also acquired many copies of the human *Alu* sequence.

Cooper's laboratory has reported similar results although there are important differences of detail⁶. Using DNA 30 kilobases or more in length (the same size as used by Weinberg's group), they obtained no transformation with DNA from MC-transformed cells of murine and avian origin or with DNA from corresponding normal cells. However, if the DNA was sheared to between 0.5 and 5 kilobases, transformation was obtained with both normal and transformed cell DNA. Cooper and his colleagues examined only one line of MC-transformed mouse cells so their failure to detect transformation with high-molecular-weight DNA does not conflict with the data from Weinberg's group, who found that only one-third of such lines scored positive in the transfection assay. The frequency at which Cooper's group observed transformation

Peter W.J. Rigby is a member of the Cancer Research Campaign Eukaryotic Molecular Genetics Research Group, Imperial College of Science and Technology, London.

with normal cell DNA, even after sonication, was extremely low and Shih and colleagues, observing such transformation, in fact at a higher frequency, regarded these data as of dubious significance. When Cooper *et al.* used DNA extracted from their primary transformants in a second cycle of transfection, they again observed transformation but at a much higher frequency (0.1–0.2 FFU μg^{-1}), comparable with that obtained by Shih *et al.* for the lines that they scored as positive. Cooper *et al.* have also performed virus rescue experiments which argue against a role for retroviruses and have used restriction enzyme digestion to show that there are at least two normal chicken cell genes capable of inducing transformation.

The following general picture emerges from these results. DNA from some chemically induced and spontaneous tumours will induce transformation even in high-molecular-weight form. In these cases the transforming gene is dominant and the genetic lesion causing transformation is likely to be *cis*-acting, for example an up mutation in the promoter controlling transcription of the gene. It is not possible to conclude that the gene detected is the primary target of the chemical carcinogen. In many other cases, however, the transformed phenotype is not transferable and the genetic lesion may therefore be *trans*-acting, for example a defect in a diffusible regulatory molecule. In these cases transformation may be induced if the DNA is sonicated to a small size before transfection. The model favoured by Cooper's group is that the sonication separates the transforming gene from regulatory sequences that normally control its expression. In the primary transfection, which is very inefficient, transformation occurs if the gene integrates close to a strong promoter. In a secondary transfection a high efficiency is obtained because the lesion is now *cis*-dominant.

This model depends on dosage. It says that genes which during normal expression are not deleterious, and may even be vital for normal growth and development, can be oncogenic when joined to regulatory elements which increase their expression. This model has received direct support from experiments by Vande Woude and his colleagues⁷. They have cloned a proto-oncogene, the murine cellular *src* gene, from normal mouse cells and shown that it does not transform when transfected into cells. If, however, the cellular *src* gene is joined to sequences from the 5' end of murine leukaemia virus which contain the strong viral promoter, the resulting recombinant DNA molecules transform efficiently.

Many questions remain. Clarification of the ability of normal cell DNA to transform and the effect of DNA size on transformation frequency is particularly important. While Weinberg's group have treated their data with great caution, they

are amenable to the interpretation favoured by Cooper and his co-workers, and one eagerly awaits the results of more detailed comparative studies. A second question is how many transforming genes are there and what is their relationship to the genes found in transforming retroviruses? The transfection assay should allow the detection of many transforming genes and Rapp and Todaro⁸ have published experiments suggesting that it will be possible to rescue genes capable of inducing a wide variety of tumours using retroviruses as vectors. The demonstration that the malignant phenotype of chemically transformed cells can be transferred by DNA transfection provides a ready assay which will allow the corresponding genes to be cloned. The availability in cloned form of genes

involved in chemical carcinogenesis will surely lead to enormous advances in our understanding of this process. Of critical importance is the demonstration that human transforming genes can be detected by transfection because the failure to isolate human transforming retroviruses has meant that we remain totally ignorant of the properties of the proto-oncogenes which are surely present in the human genome. [1]

1. Cold Spring Harb. Symp. quant. Biol. 44 (1980).
2. Hayman, J. Gen. Virol. 52, 1 (1981).
3. Bishop, J.M. Cell 23, 5 (1981).
4. Shih, C. *et al.* Proc. natn. Acad. Sci. U.S.A. 76, 5714 (1979).
5. Shilo, B.Z. & Weinberg, R.A. Nature 289, 607 (1981).
6. Cooper *et al.* Nature 284, 418 (1980).
7. Oskarsson *et al.* Science 207, 1222, (1980).
8. Rapp & Todaro Proc. natn. Acad. Sci. U.S.A. 77, 624 (1980).

Proline and protein procrastination

from Roger H. Pain

EVER SINCE it was realized that a polypeptide could not afford the time to explore all possible configurations before settling in the unique native conformation, the main thrust of work in the field of protein folding has been concerned with understanding how folding can take place so rapidly. The sequences of proteins are designed not only to ensure appropriate stability and biological function but also, and of equal importance, to direct the folding reaction along pathways which shortcut the majority of configurations and allow the native conformation to be reached in well under a minute. The broad outlines of how this happens are now fairly clear.

It is therefore strange, at first sight, to find that a mechanism currently the object of a substantial amount of effort is one whereby this highly sophisticated physico-chemical process is actually delayed by the molecular eccentricity of a few proline residues (see Freedman *Nature* 279, 756; 1979). The biphasic kinetics of refolding exhibited by many proteins have been explained by the general model



where U_S is an unfolded conformation which folds to the native state, N, with relatively slow kinetics because it has to undergo a prior rate-limiting change into the fast refolding form U_F . U_F contains prolines in the *cis* or *trans* isomers which

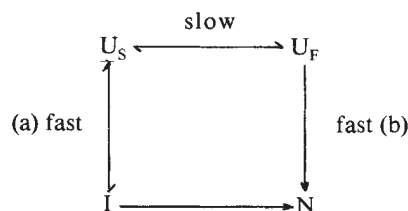
occur in the state N, but U_S is a form in which a proportion of non-native isomers have arisen by slow *cis-trans* isomerization. Thus, the kinetics of refolding are modified by a transformation which is itself almost spectroscopically invisible.

The evidence for this explanation is, in the words of Jullien and Baldwin (*J. molec. Biol.* 144; 1981), "strong but not yet complete". It has been supported by the key double-jump type of experiment. If state N is perturbed to state U_F by thermal, acid or solvent denaturation (jump 1), the kinetics of the refolding to N after removal of the perturbation (jump 2) will depend on the time allowed for the $U_S \rightleftharpoons U_F$ equilibrium to become established. It is found that the slow phase becomes more pronounced the longer the time interval between the first and second jumps and that the activation energy is similar to that for proline isomerization. Furthermore, strong acid, which catalyses proline isomerization, also increases the rate of equilibration between U_F and U_S .

The above equilibrium is the simple model initially put forward in 1975 by Brandts, Halverson and Brennan (*Biochemistry* 14, 4953). However, it has always seemed intuitively possible that U_S could fold to a compact configuration, still containing the 'wrong' proline isomer, but then reorganize itself on subsequent isomerization to state N. Baldwin and his collaborators have shown that for ribonuclease this is true and, indeed, that the isomerization process can be speeded up in this folded state by as much as

Roger H. Pain is in the Department of Biochemistry at the University of Newcastle upon Tyne.

50-fold. The more general basic model for refolding is now



where I represents one, or a sequence, of intermediate states with improper prolines. Levitt has drawn attention to the consequence inherent in this model that the refolding route taken will be a function of the difference in conformational energy between the native state N and the similar though perturbed conformation containing the wrong proline isomer (*J. molec. Biol.* **144**; 1981). The higher the strain associated with the incorporation of incorrect proline isomers into the native conformation, the greater the relative importance of pathway (b) and the more the system will approximate to Brandts' original model. At the other extreme, if the difference in energy is low, pathway (a) will predominate and isomerization only be detectable insofar as the folded state I differs spectroscopically or in stability from state N. State I will be expected to be significantly populated under conditions favouring the 'native' state.

Levitt, on the basis of conformational enthalpy calculations, postulates three levels of strain energy resulting from the insertion of incorrect proline isomers, types I, II and III corresponding to energy differences of approximately 1, 10 and 30 kcal mol⁻¹. Types III and I correspond, respectively, to the two cases cited above, while in type II the folded state containing an incorrect proline would be relatively stable but significantly less so than the native state.

Jullien and Baldwin have taken up this classification and have suggested kinetic and environmental criteria whereby the three types may in principle be recognized experimentally. On the basis of spectroscopic and urea gel electrophoresis (Creighton *J. molec. Biol.* **137**, 61; 1980) investigations on about ten proteins, they find definite evidence for proteins with type II prolines only, although the existence of type I prolines cannot be ruled out. Thus, in any system some molecules refold fast, others slowly. Generalizations about globular proteins are notoriously prone to contradiction, however, and investigation of a wider range of proteins may reveal different behaviour.

Levitt's main contribution is to calculate the strain energies for each of the four prolines placed, as the incorrect isomer, in bovine pancreatic trypsin inhibitor (BPTI). The effect on the protein is shown to involve mainly local and relatively small shifts from the native conformation. He finds that the destabilization of the native

structure is 1 kcal mol⁻¹ for Pro 13 (therefore type I), 11 kcal mol⁻¹ for Pro 2 and Pro 9 (type II) and 33 kcal mol⁻¹ for Pro 8 (type III). On the basis of these assignments, one can calculate the proportion of BPTI molecules which will refold rapidly and Levitt obtains a value of 0.77. Jullien and Baldwin, who carry out a detailed experimental investigation of the refolding of a slightly modified BPTI, find that the relative amplitude of the fast folding U_F → N reaction is 75 per cent, in good agreement with the prediction. They also find one slow folding reaction which satisfies the criteria for a type II residue together with preliminary evidence strongly suggestive of one or two type I residues. Type III behaviour has not been found, although this may be due to greater conformational mobility in the modified BPTI which they study. The 'catalysis' of isomerization found experimentally has been studied by Levitt who calculates that states intermediate between *trans* and *cis*, with the peptide bond angle $\omega \sim 90^\circ$, have lower energies than the strain energies in

the *cis* form. The protein seems to be able to accelerate the rate of isomerization by applying a couple to the ω angle.

Taken together, these two different approaches provide more circumstantial evidence for proline isomerization as the explanation for the biphasic refolding of proteins. If the fast and slow phases do, in fact, correspond to the sequential U_S → I and I → N reactions where I is a compact configuration, and if most proteins turn out to fold as type II or type I, the proline quirk affects the exquisite efficiency of the folding process less than might have been initially feared. The polypeptide chain still collapses rapidly, as suggested by Creighton, to a compact form within which the prolines can sort themselves out at leisure. The collapse may be the most important feature in that it reduces the possibilities for entanglement between elements of different domains. Presumably the conformational requirements for proline residues in the folded protein outweigh the necessity for an overall rapid folding process. □

Defining a lectin

from Jan Kocourek and Václav Hořejší

Recent correspondence (see *Nature* **285**, 66; 1980) demonstrates that there is an urgent need to formulate a definition of lectins which clearly delineates the meaning of the term. However, Goldstein *et al.*'s proposed definition of a lectin as "a sugar-binding protein or glycoprotein of non-immune origin which agglutinates cells and/or precipitates glycoconjugates" seems to disregard some important points. In particular, because the role of lectins in living objects is unclear, it does not seem appropriate to define lectins by their *in vitro* biological activities. The agglutination of cells — a complex and poorly understood phenomenon — may, in the majority of cases, be a fortuitous effect of lectins. The definition of lectins should be based primarily on their most important physicochemical properties, that is, on their interaction with carbohydrates.

The inclusion into the definition of some *in vitro* activities of lectins, like agglutination of cells and precipitation of glycoconjugates, would exclude those lectins which have only one binding site and/or show some toxic or hormone-like activities. This is especially true in the case of the toxins of *Ricinus communis*

and *Abrus precatorius* which are structurally related to the 'agglutinins' and occur in the same sources so that their common evolutionary origin is obvious. Recent research shows that many of these agglutinins (including that of *R. communis*) may exert very similar properties to their toxic counterparts, if acting in cell-free systems (Barbieri *et al. Biochem. J.* **182**, 633; 1979).

As long as we consider lectins as recognition molecules and assume that they mediate information in various (but still largely unknown) physiological reactions, their definition should be based on only three distinct features which they all have in common. Lectins

- (1) have at least one binding site which can bind free or glycosidically linked carbohydrates: this interaction can be inhibited by low-molecular-weight sugars;
- (2) are not synthesized due to an immune response and do not have immunoglobulin structure;
- (3) do not show enzymatic activities towards carbohydrates to which they bind.

We thus suggest the following definition: lectins are sugar-binding proteins or glycoproteins of non-immune origin which are devoid of enzymatic activity towards sugars to which they bind and do not require free glycosidic hydroxyl groups on these sugars for their binding.

Jan Kocourek is in the Department of Biochemistry, Charles University, Prague and Václav Hořejší is in the Institute of Molecular Genetics, Czechoslovakia Academy of Sciences, Prague.

Chondritic meteorites: a changed and changing view

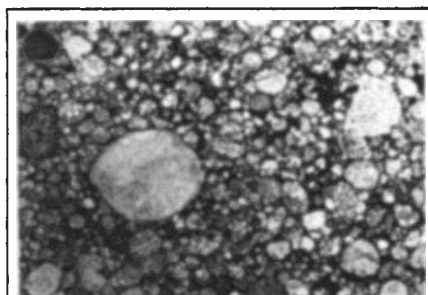
from R. T. Dodd

Two decades of intensive study of meteorites has drastically changed our view of their role in Solar System history. What R.A. Daly regarded in 1943 as fragments of a shattered tenth planet¹, we now recognize as samples of at least 20 (ref. 2) and perhaps more than 70 parent bodies³, most or all of which were no larger than Ceres, the largest modern asteroid (with a radius of 385 km)^{2,4}. Some of the meteorite parent bodies experienced partial melting and fractional crystallization to yield the diverse rocks that we know as iron, stony-iron and achondritic stony meteorites. Others escaped such differentiation and yield samples of rocks — chondritic meteorites — that formed during or very shortly after condensation of the solar nebula and underwent little physical and less chemical alteration thereafter. While the differentiated meteorites provide valuable insights into planetary evolution⁵, chondrites contain a unique record of the earliest stages in Solar System history, and some of them contain isotopically bizarre materials that may predate formation of the solar nebula and/or come from beyond it⁶. All meteorites are important, but the chondrites alone among accessible materials are survivors of the first steps in the 4.55 Gyr progression from interstellar gas to modern man.

Chondrites are more or less recrystallized agglomerate rocks (see the figure) whose bulk chemical compositions (ignoring H, He and other gases) resemble the composition of the Sun. Though all chondrites are, in this sense, chemically primitive, they show discontinuous variations of oxidation state and ratios of major non-volatile elements that divide them into three chemical classes and at least eight smaller chemical groups (see the table). There is general agreement that these variations reflect processing of chondritic material before or during its accumulation in at least eight parent bodies, though the nature of that processing remains unclear^{4,7,8}. There is also a consensus that the recrystallization evident in most ordinary and enstatite and some carbonaceous chondrites signifies the thermal metamorphism in^{4,9} or on^{8,10,11} the

surfaces of the parent bodies; the specific mechanism is again unclear⁴.

Much discussion of chondrites centres on the origin of chondrules, the millimetre-sized, drop-formed or irregular bits of rock from which these meteorites take their name (see the figure). Chondrules are present in most carbonaceous and all enstatite and ordinary chondrites, and their origin is highly controversial. Of the



Photomicrograph, in transmitted light, of the Manych ordinary chondrite. Chondrules, which comprise almost 90 per cent by volume of this meteorite, consist chiefly of olivine, pyroxenes and residual glass. The field of view is 16 mm wide.

many processes invoked to explain them — condensation of droplets from nebular gas, fusion of nebular dust by various means, shock melting of dust or rock on parent body surfaces, explosive volcanism and metamorphism — only the last two have no modern proponents, and only the last is clearly inconsistent with the properties of chondrules.

Debate over the origin of chondrules remains spirited, but new analytical methods (electron microprobe, neutron activation and isotope dilution analysis) have provided analyses of individual chondrules that place tight constraints on models. Chemical trends among chondrules in carbonaceous chondrites strongly suggest that these objects, like the refractory inclusions in such meteorites, formed in the condensing solar nebula, by condensation of droplets and/or fusion of solid condensates¹². On the other hand, chemical variations among chondrules in

ordinary chondrites suggest that they post-date condensation and formed through shock melting associated with the growth of parent bodies¹³. It thus appears necessary to appeal to different processes and/or times of formation⁴ to explain chondrules in carbonaceous and ordinary chondrites.

Chondrules in enstatite chondrites have received little attention to date, largely because of the perception that all such meteorites are metamorphic rocks¹⁴. Two recent papers on such chondrules^{15,16} are therefore welcome. They suggest that metamorphism is a less severe problem than believed, and that the chondrules in enstatite chondrites are remarkably complex objects that may require yet another explanation.

Leitch and Smith showed that chondrules in the Indarch chondrite contain two types of clinoenstatite (essentially MgSiO_3)¹⁵ and two types of olivine (essentially Mg_2SiO_4)¹⁶ that differ in minor element contents and cathodoluminescence colour. That these and similar phases in the Kota-Kota chondrite¹⁶ occur together but do not intergrade makes it unlikely that either meteorite experienced significant metamorphic alteration. In the standard classification of chondrites by composition and degree of recrystallization¹⁴, they probably belong to type E3 rather than E4.

Leitch and Smith (ref. 15 and see this issue of *Nature* p.228) observed that the distribution of the two pyroxenes and two olivines is very complex — some chondrules contain all four minerals — and hard to reconcile with a direct connection between chondrule formation and nebular condensation. They suggest that enstatite chondrites bring together materials from at least two nebular sources, whose history included condensation, formation of planetesimals, collisional destruction of these objects and accumulation of the resulting debris.

Many aspects of this scenario need further study, as its authors admit. It is important, for example, to characterize the fine material between crystals in chondrules, which Leitch and Smith assume to be dust and vapour deposits. This writer is also troubled by their explanation of droplet chondrules, which tend to contain either red- or blue-luminescing pyroxene (or orange or blue olivine) and are thought to sample different planetesimals. It is not clear why liquids generated by collisions should sample only one of the colliding objects.

R. T. Dodd is Professor of Mineralogy in the Department of Earth and Space Sciences, State University of New York at Stony Brook.

Abundances and chemical properties of chondrite classes

Class	Groups	Falls	Fe/Si	Atomic Fe ^o /Fe	Mg/Si
Carbonaceous	>4	33	1.6–1.8	0.0–0.1	0.90–0.92
Ordinary	3	314	1.2–1.6	<0.1–0.6	0.80–0.83
Enstatite	>1	11	1.1–2.0	0.7–0.8	0.63–0.68

Abundance data from ref. 8; approximate ranges of ratios from ref. 4.

The Leitch-Smith model for chondrule formation is at present skeletal, but their demonstration that essentially pristine enstatite chondrites exist is very important. It encourages those (for example ref.16) who are attempting to study chemical variations among chondrules in these meteorites. It should also encourage other workers to move into this neglected corner of chondrite research.

Evidence that the chondrules in carbonaceous, ordinary and enstatite chondrites tell quite different stories illustrates very well the marked change in view alluded to at the beginning of this article. Daly¹ interpreted chondrites as samples of a planetary mantle and thus, by implication, chemically and physically differentiated materials. Had he proved to be correct, we might now know more than we do about the interior of our planet, but we would surely know much less about its origin. A shattered tenth planet is emotionally appealing, but it would tell us much less about the Solar System's history than 20 or

more smaller objects caught at various stages of internal evolution. That the meteorite parent bodies were numerous and varied might disappoint Daly, but it is very good news for planetary scientists. □

1. Daly, R.A. *Bull. geol. Soc. Am.* **54**, 401 (1943).
2. Wasson, J.T. & Wetherill, G.W. in *Asteroids* (ed. T. Gehrels) 926 (1979).
3. Scott, E.R.D. *Geochim. cosmochim. Acta* **42**, 1447 (1978).
4. Dodd, R.T. *Meteorites: A Petrologic-Chemical Synthesis* (1981).
5. Stolper, E., McSweeney, H.Y. Jr & Hays, J.F. *Geochim. cosmochim. Acta* **43**, 589 (1979).
6. Clayton, R.N. *A. Rev. nucl. particle Phys.* **28**, 501 (1978).
7. Larimer, J.W. & Anders, E. *Geochim. cosmochim. Acta* **34**, 367 (1970).
8. Wasson, J.T. *Meteorites, Classification and Properties* (1974).
9. Dodd, R.T. *Geochim. cosmochim. Acta* **33**, 161 (1969).
10. Heyse, J.V. *Earth planet Sci. Lett.* **40**, 365 (1978).
11. Hutchison, R., Bevan, A.W.R., Agrell, S.O. & Ashworth, J.R. *Nature* **280**, 116 (1979).
12. McSweeney, H.Y. Jr *Geochim. cosmochim. Acta* **41**, 1843 (1977).
13. Dodd, R.T. *Earth planet Sci. Lett.* **40**, 71 (1978).
14. Van Schmus, W.R. & Wood, J.A. *Geochim. cosmochim. Acta* **31**, 747 (1967).
15. Leitch, C.A. & Smith, J.V. *Nature* **283**, 60 (1980).
16. Denham, M. Personal communication (1981).

Why is the moon slowing down?

from David W. Hughes

THE Moon orbits the Earth every month and on average moves 13.2° per day across the sky. As long ago as 1695 Edmond Halley (*Phil. trans. R. Soc.* **19**, 1960; 1695) suspected that the Moon was slowing down in its orbit. Now, that the effect has been confirmed, two main causes have been proposed: tidal dissipation and variation in the value of G , the universal 'constant' of gravitation. Measurements of this acceleration have been reviewed and interpreted by F.R. Stephenson of the Department of Geophysics, Liverpool University in a recent edition of the *Journal of the British Astronomical Association* (**91**, 136; 1981).

Tidal friction retards the Earth's rotation. The length of the day is thus slowly increasing at a rate of about one second in 50,000 years. As an autonomous system conserves its angular momentum, the angular momentum lost by the Earth is gained by the Sun and Moon. The Sun is so massive that the reciprocal effect on it is negligible but the Moon does, however, gain a significant amount of angular momentum at the expense of the Earth. This causes our satellite to drift away at a rate of about 4 cm per year.

The acceleration of the Moon can be calculated in two independent ways. The first relies on the analysis of ancient eclipse

observations which extend back over 2,000 years. The difference between the predicted lunar position, if we neglect acceleration, and the observed position which obviously is affected by the acceleration, increases as the square of the elapsed time. Over 2,000 years this difference is as large as two or three degrees. The most recent result obtained using this technique is -30 ± 3 arc sec century⁻². An indirect method based on the observations of the transits of Mercury since 1743 produced a result of -26 ± 2 arc sec century⁻².

The second approach is through the study of oceanic tides. Recent research has shown that the sea level has not varied by more than a few metres over the last three millennia and there is no evidence of any significant change in the lunar tidal friction over the past 2,500 years. It has also been found that the vast open oceans are the major energy sink and not, as previously thought, shallow regions such as the Bering Sea. Calculations of the mean rate of work done by the Moon on the ocean surfaces resulted in an acceleration of -30.6 ± 3.1 arc sec century⁻². Ocean tides perturb the orbits of Earth satellites and an analysis of these perturbations lead to a result of -27.4 ± 3 arc sec century⁻².

The recession rate $\dot{r}$ is related to the accel-

eration $\dot{n}$ and the rate of change of the constant of gravitation $\dot{G}$ by the following equation:

$$\frac{\dot{r}}{r} = -\frac{2\dot{n}}{3n} + \frac{1}{3}\frac{\dot{G}}{G}$$

If $\dot{n}$ is measured and $\dot{G}$ is assumed to be zero, $\dot{r}$ can be easily calculated. The present value for $\dot{r}$ is 4.0 ± 0.3 cm yr⁻¹. The Moon was once much closer to the Earth than it is now. G.H. Darwin (*Scientific Papers* **2**, 208, Cambridge, 1908) originally extended this deduction to claim that the Moon may have originated as part of the Earth. The recession rate is inversely proportional to the sixth power of the distance. The assumption that the present rate of energy dissipation has always been active would bring the Moon catastrophically close to Earth less than 1.5×10^9 yr ago. The Earth is three times older than this. Here, to quote W.H. Munk (*Quarterly Journal of the Royal Astronomical Society* **9**, 352; 1968), "we are extrapolating from a few thousand years to a few thousand million years, always a dangerous undertaking, and a harrowing one when the very process is in doubt". Fortunately it has been shown that even moderate changes in the shape of oceans and adjacent seas may materially affect the tidal dissipation and hence the recession rate. The astronomical approach thus seems to be of little help in solving the problem of exactly where the Moon was in the geological past.

The difference between the acceleration measured as a function of Ephemeris Time and the acceleration measured as a function of Atomic Time yields the non-tidal component and consequently the rate of change of G . Some 7,000 observations of lunar occultations of stars have been used by Van Flandern to find an Atomic Time value of the acceleration of -36 ± 5 arc sec century⁻². Laser reflectors have been also placed on the Moon by Apollo astronauts and have been used to give a lunar acceleration with an Atomic Time value of -23.8 ± 4 arc sec century⁻².

Stephenson carefully combines the Ephemeris Time and Atomic Time values of the acceleration. There seems to be considerable overlap between the two data sets and he concludes that the present observational evidence is quite consistent with the interpretation that $\dot{G}$ is zero and the universal constant of gravitation is not varying.

What will be the end result of lunar recession? The day (the Earth's spin period) and the month (the lunar orbital period) are both increasing, the day faster than the month. Eventually it will take Earth fifty-five of our present days to spin once. It will also take the Moon fifty-five days to orbit. Day and month will be equal and the same part of the Earth's surface will always face the Moon. □

David W. Hughes is in the Department of Physics, University of Sheffield.

Intergenic conversion and reiterated genes

from Richard Egel

GENE conversion between reiterated genes is currently gaining recognition as a novel means for exchange in the genome. The term gene conversion stands for the replacement of a particular DNA sequence by a different one. This process is usually initiated by the formation of hybrid DNA between two partially complementary single strands. In general, they belong to two double-stranded molecules. The following discussion, however, deals particularly with the special case when both originate from slightly different parts of the same chromosome.

Until recently, gene conversion has been the esoteric refuge of a few fungal-geneticists who studied the hidden secrets of recombination from aberrant tetrad configurations, joined only by a few determined drosophilists who examined 'half tetrads' of 'attached X' chromosomes, in which the two homologues are connected by a common centromere. These studies suggest that heteroduplex or hybrid DNA is a local intermediate in recombination events of any sort. In a single meiosis, any pair of allelic markers should segregate 2:2 unless they are drawn into a recombination event, in which case various deviations from the regular 2:2 distribution may and do occur. In fungi with four spores per tetrad, 3:1 events are the easiest to detect. In other fungi with eight spores per tetrad, 6:2 and 5:3 events were the first aberrant classes analysed. Originally, such a skewed segregation pattern, where one of the markers had fallen into minority, was classified as 'gene conversion', but it is now common practice to call the formation of hybrid DNA a conversion event, irrespective of whether the final segregation pattern becomes skewed or remains symmetrical.

To understand recombination, one must focus on a single intermediate — the piece of hybrid DNA where two single strands from different molecules have reannealed (Fig. 1). In general, such a reannealing will not lead to a perfect match as any difference in base sequence will reveal itself as a misalignment, comprising single bases or larger loops of noncomplementarity (a). Such discontinuities are either left alone until they are resolved by replication (b), or are erased before replication by repair synthesis (c), using one or the other strand as the conserved template.

If one now considers the adjoining stretches of homoduplex DNA on either side of the hybrid, there are two basic possibilities (Fig. 2): the flanking sequences can originate from the same molecule (a), or they may come from the two molecules that contributed to the

Richard Egel is a professor at the Institute of Genetics, University of Copenhagen.

Hybrid DNA as an intermediate in recombination

Fig. 1 A stretch of hybrid DNA (a), including a noncomplementary region, can be replicated directly (b) or first corrected by repair synthesis (c).

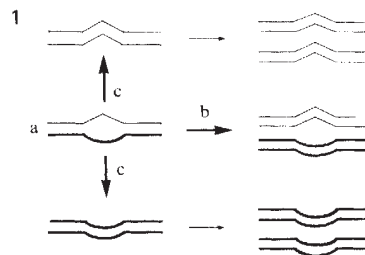


Fig. 2 The hybrid region can be part of a continuous molecule (a) or recombine two different ones (b).

Fig. 3 The 'Aviemoore model' of hybrid DNA formation between two molecules¹. Newly synthesized DNA; ——— degraded strand.

hybrid (b). In the first case a limited piece of single-stranded DNA has invaded another duplex without disrupting the overall arrangement of that molecule. In the second, the outer parts have assumed a recombinant configuration.

Usually both the DNA molecules engaging in recombination come off intact. This conservation can be accounted for by the following dynamic model¹, which also takes care of other observations (Fig. 3). At special sites, one of the strands is opened (O), and one of the single-stranded ends begins to invade the partner molecule. The leaving strand is peeled off further, and the gap in its wake is filled by DNA synthesis, using the remaining strand as a template. It may, in fact, be this synthesis that actually drives the reaction. In the receptor molecule, the strand displaced by the invader seems to be degraded initially, but at any moment (S), it can likewise start to invade the partner molecule, whereafter the further formation of hybrid DNA will be symmetrical. This two-phase scenario with a switch from asymmetrical to symmetrical formation of hybrid DNA has been verified experimentally in the fungus *Ascobolus*^{2,3}.

At the final stage (H), the connected molecules are known as the 'Holliday structure' or a 'half chiasma'. The junction is relatively free to move back and forth by the concerted rotation of the four double-stranded arms. Eventually, however, this connected structure needs to be resolved. In meiosis, where recombination is generally thought to occur between two of the four chromatids available after replication, the Holliday structure must be

broken by the cutting of two equivalent strands — the two 'inner' ones or the two 'outer' ones, as placed in the drawing (H). These two alternatives may well assume a less asymmetrical configuration when viewed in three dimensions. In mitotic recombination, however, when hybrid DNA appears to be formed at the two-chromatid stage, resolution is more likely to result from replication through the Holliday structure⁴, rather than by strand cleavage before replication.

A priori, there is no reason why the recombination processes outlined above cannot occur within families of reiterated genes, such as the rRNA, globin or antibody genes. Hybrid DNA can still be formed as long as there is sufficient sequence homology. As the flanking regions may not be homologous, however, their rearrangement by crossing-over will not, in general, be tolerated as this might lead to a change in the arrangement of genes. However, the invading strand can also be disconnected from its source, so there could be recombination between the reiterated coding sequences alone. In this way the overall arrangement of the receptor chromatid would remain the same. This type of information transfer between related but different loci has recently been documented in various organisms.

That recombination essentially depends on local DNA homology is widely accepted. For example, when the *rosy* gene of *Drosophila* is placed inside a heterozygous inversion, the frequency of gene conversion is not changed⁵, whereas crossing-over in the inverted region is

deleterious, generating a dicentric chromosome together with an acentric fragment.

Duplicated genes are often located in a tandem arrangement. This situation can now be simulated experimentally by gene-cloning techniques. Klein and Petes⁶ selected a yeast transformant, in which the wild-type *leu2*⁺ gene, together with a bacterial vector, was integrated in the vicinity of the chromosomal *leu2* gene, which was marked by two frameshift mutations. This strain was then crossed with a *leu2*⁺ partner. Hence, one of the chromosomes of the diploid carried two *leu2* genes in tandem, only separated by the vector DNA. In the absence of recombination, all four spores of a tetrad should receive a copy of the *leu2*⁺ allele and grow without leucine. Occasionally, however, tetrads were observed in which one spore was auxotrophic, due to one or both of the original frameshift mutations. When these aberrant tetrads (12 of 306) were analysed further, the largest fraction (6) was due to conversion between the duplicated genes. Four cases were due to conversion between non-sister chromatids, and only one was due to crossing-over. One tetrad was apparently caused by a multiple event, but it could also be explained by early conversion, if hybrid DNA was formed before the premeiotic S phase — similar to what is usually observed for mitotic recombination. On the basis of these results, the authors conclude that the cells can use gene conversion to maintain sequence homogeneity in a family of reiterated genes. For this purpose, conversion is more efficient than crossing-over at staggered places in a tandem series, the rectification mechanism previously proposed⁷ because with conversion the overall number of genes will be maintained.

Circumstantial evidence for the occurrence of similar events outside the laboratory has recently been detected in the DNA sequences of three human globin genes⁸, two of which were reiterated on the same chromosome with the third allelic to one of the others. As is usual for globin genes, the expressed parts of these genes are interrupted by several introns, and it is one of these that apparently has been subject to intergenic conversion. This conclusion was based on the observation that two equivalent introns of the reiterated genes contained a segment more similar to each other than to the allelic counterpart. Interestingly, this segment was located adjacent to a presumptive hot spot of recombination.

Working with quite a different system, Munz and Leupold⁹ reported genetic evidence for the specific exchange of information between reiterated tRNA genes in fission yeast. They studied the reappearance of suppressor-active alleles that had been inactivated by secondary mutations. This exchange was also most easily interpreted as intergenic conversion.

Intergenic conversion has also been

applied to an analogous phenomenon — mating-type switching in homothallic yeast strains¹⁰⁻¹². It appears that a certain piece of DNA is transferred from a silent position to an expressible site, where it replaces the piece of information previously residing there. This switches the functional mating type to the opposite type. The information transfer is highly asymmetrical in the sense that the source segment is usually left unaffected, and the target segment to be displaced usually discarded. The organism has thus adopted a conversion mechanism for the purpose of gene control, and it has done so by channelling the course into a unique direction.

What other systems are likely to benefit from similar explanations? Gene conversion may perhaps play a role in the generation of antibody diversity. Reciprocal recombination can apparently explain a host of data from V-C joining to class switches. However, the experts still appear divided over whether hypervariable regions can recombine as mini-genes¹³. Their independent incorporation into the framework of the V genes would certainly increase considerably the combinatorial

possibilities. This idea, alas, was badly hit by the discovery of a full set of hypervariable regions in the first V gene from the germ line to be sequenced. However, that blow need not be as fatal as it seems, because only the downright insertion of new sequences is immediately excluded, and not the substitution of one sequence for another of equivalent length. That is precisely what conversion has to offer. So there is still room for mini-genes to jump about, and the search for more evidence need not be called off just yet. □

1. Meselson, M. & Radding, C.M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 358 (1975).
2. Paquette, N. & Rossignol, J.L. *Molec. gen. Genet.* **163**, 313 (1978).
3. Rossignol, J.L. & Haedens, V. *Curr. Genet.* **1**, 185 (1980).
4. Esposito, M.S. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4436 (1978).
5. Chovnick, A. *Genetics* **75**, 123 (1973).
6. Klein, H.L. & Petes, T.D. *Nature* **289**, 144 (1981).
7. Petes, T.D. *Cell* **19**, 765 (1980).
8. Slightom, J.L., Blechl, A.L. & Smithies, O. *Cell* **21**, 627 (1980).
9. Munz, P. & Leupold, U. *Alfred Benzon Symp.*, 16 (in the press).
10. Klar, A.J.S., Strathern, J.N., Broach, J.R. & Hicks, J.B. *Nature* **289**, 239 (1981).
11. Klar, A.J.S., McIndoo, J., Strathern, J.N. & Hicks, J.B. *Cell* **22**, 291 (1980).
12. Leupold, U. *Nature, News and Views* **283**, 811 (1980).
13. Robertson, M. *Nature, News and Views* **287**, 390 (1980).



100 years ago

In a paper "presented to both Houses of Parliament" the subject of "oleomargarine" as manufactured in the United States is discussed. This substance is made from beef suet by disintegrating in warm water, passing through a fine sieve, melting at 120° F., settling, draining off the oil, and allowing to solidify. During the year ending June 1880, 18,833,330 lbs. of oleomargarine were exported from New York, the greater part going to Holland. The manufacture and sale of this substance is strongly condemned by many butter merchants, and as strongly recommended by various well-known American chemists. Analyses given in the report show very small differences between oleomargarine and natural butter, except in the particular of soluble fats, of which oleomargarine contains considerably less than natural butter.

Cave Animals and Multiple Centres of Species

The readers of Semper's "Existenzbedingungen der Thiere," now translated into English, will find (vol.ii, p.268 of the German edition) an interesting discussion of the question of monophyletic or polyphyletic evolution of species, the author decidedly inclining to the latter hypothesis. Considering that at the root of the manifold and difficult problems here involved, there is the relatively simple one of single or multiple centres of each species in a biographical sense, I take leave to ask the following question, hoping for an

answer from among your readers versed in these matters.

To me it seems impossible to maintain the single centres of species in a strict and definite sense without also maintaining the single progenitor of each species, which latter view, formerly considered as a necessary assumption, has been given up by Mr. Darwin in Chapter IV. of the later editions of "Origin of Species" (5th ed. p.103, 104). Of course the acceptance of single centres, in the sense of more or less restricted areas of origination, may remain valid for the vast majority of species.

Now, I have sometimes thought that there might be a test for the possibility of multiple centres, which, eventually, would amount almost to an experimental demonstration — namely: *whether there are cases of the same species of blind animals occurring in different caves distant from and without subterranean communication with each other?* Should such cases occur it would be most improbable that the animals in question had been transported from one cave to the other in the modified state, and most probable that they had been independently evolved in each cave from identical species which entered it from without. I formerly noted one instance perhaps in point, viz. a statement of Prof. Cope's (*Nature*, vol.vii, p.11) that "the blind fish of the Wyandotte Cave is the same as that of the Mammoth, the *Amblyopsis speculatus*, Dekay," but I am not aware whether subterranean communication is, or has been, impossible in this instance. Perhaps more decisive cases have become known of late?

Freiburg im Breisgau, March 4

D. WETTERHAN

from *Nature* **23**, 17 March, 458, 469, 1881.

COMPUTERS IN THE NEW LABORATORY

Where will the revolution end?

Computers have come to stay. And the uses of computing machinery are by no means confined to the first and more obvious applications, the solution of problems in large-scale arithmetic. The pages that follow are intended to point to some of the more conspicuous ways in which computing machinery is being used for the solution of problems in scientific research which are not mere computation. What follows is not, however, a survey of this now vast field but rather a pointer to some of the trends that have become apparent in the past few years. There is every prospect that, in the years ahead, the scientific uses of computing machinery will grow still further. Where will it all end? And what will be the consequences for the practice of scientific research?

One of the most conspicuous developments in the past few years has been the use of computing machinery in the handling of data, but the passage of time serves only to emphasize how much remains to be done. In many circumstances, there is no alternative to computerized handling. It is unthinkable that the data collected by, say, the Einstein X-ray satellite could be handled in any other way. The same is true of the vast amounts of data collected by detection equipment in high-energy physics, some of which yield information in a thousand or so parallel channels in intervals of time no greater than a microsecond. In these respects, computers have literally made possible some kinds of experiments. But data-handling techniques also enhance the value of data. The Einstein satellite is again a striking illustration of what can be done (see page 197). If the detectors (by their nature) record the time at which X-ray photons arrive, or their energy, why not design the data-handling system to exploit this extra information? In this respect, computers can wring more out of an experiment, or a series of observations, than is possible with more traditional techniques of data handling. It is natural that astronomers of all kinds should be alive to the possibilities with which they are now presented. Others will no doubt follow suit.

Another of the important trends of the past few years (hardly referred to in what follows) is the use of computing machinery for simulation. Some of the most ambitious essays of this kind are the climatic models that have been developed at half a dozen centres around the world. The objective is to represent within the circuitry of a computer the physical processes occurring in the real atmosphere. Climatic models are notorious in their demands of the computer hardware, while the development of climatic programs is hampered not only by the complexity of the problems but by ignorance of, or at least uncertainty about, some aspects of the underlying processes. Even so, climatic models have brought valuable understanding to some important aspects of the problem of climatic change. They have surprisingly been less effective at the identification of the physical problems that must be better understood before reliable models can be designed.

Already there have been some striking and successful applications of simulation in the study of the nervous system for example. In other fields, however, simulation by computer model has lived up fully to its purpose, to demonstrate the validity of a theory or to identify its weak points, and to serve as a convenient and rapid means of telling what will happen in the real world. This is another persistent trend.

Computers that run experimental equipment are also on the march. Again, it is unthinkable that the instruments carried aloft by rockets and satellites could be operated differently, but the sophistication of the technology now in use is a quite remarkable demonstration of what has been accomplished in the past twenty

years, partly no doubt because of military interest in the same techniques. That experimental equipment in, for example, high-energy physics should be controlled in the same way is also easily understood — there is no choice. With the introduction of relatively cheap microprocessing equipment, however, other surprising applications of this kind have come to light. In some quite small laboratories, it turns out to be worthwhile to use such a piece of machinery for controlling a piece of standard equipment from which a sequence of observations is required. The benefits are certainty, speed and the saving of research assistants' time. The cost is often small.

In the nature of things, however, most excitement in the years ahead will lie in directions not now foreseen. To many people, the use of computers for carrying out routine algebraic calculations (see page 198) will be a surprise. With hindsight, of course, this development seems natural enough. But how many would have staked money or reputation on the feasibility and the success of this development ten years ago, when computer algebra was in its infancy? And who now can guess what other developments, perhaps more radical, are buried in embryonic form in the research programmes of the computer laboratories?

Many of the applications of computers in past decade have made possible tasks that otherwise could not (or would not) have been attempted. Most of the same applications have been conceived of by the people most concerned to benefit from them — researchers themselves, not the manufacturers of computer hardware. And the chances are that the burden of innovation will continue to fall on the ultimate beneficiaries. With this in prospect, there would seem to be an urgent need that researchers in all fields, not simply those in the physical sciences, should be equipped by their education to be able to appreciate the possible benefit of computer applications. Although, as the experience of recent decades suggests, the most unlikely disciplines appear to accrete their own bands of computer enthusiasts, it seems unwise that they should rely on such a random process. The implications for the education of graduate students are obvious. Some practical experience of how to put a computer to good use is essential. But it may be even more important that graduate students should be led to understand the limitations of computer machinery — and to appreciate the pitfalls of over-dependence.

Like all blessings, the new computer technology is, after all, a mixed blessing. Making computers work well can rarely (outside the computer laboratories) be an end in itself. And the pace of discovery would be constrained if too many working scientists spent too much time writing programs for machines. It may also be damaged if the seductiveness of computing machinery, and the temptation to ask too much from simple computer programs, leads laboratories into spending too great a proportion of their resources on what is, after all, merely a means to an end. □

CONTENTS

Computers and the biologist	194
Computers cheaper: Computers in physics	195
Logic circuits before computers	196
Astronomy by numbers	197
Some things old, some new	197
Algebra made mechanical	198
Plenty of programs: Algebraic computer programs	200

Computers and the biologist

Among biologists, physiologists have inevitably taken most readily to the use of computers; people brought up to know the difference between d.c. and a.c. amplifiers are naturally more inclined to give house-room in their laboratories to further electronic equipment. Yet the past few years have seen the infiltration by computers of most other kinds of research laboratories.

In all kinds of laboratories, the chief use of computers is still in the storage and analysis of data. The principles are as old as computers themselves. Data are digitized, stored within the accessible part of the computer or dumped onto magnetic tape, and analysed by whatever computer program is appropriate. In the 1960s, most large biological laboratories found it expedient to install substantial main-frame computers to provide their members with a data-processing service. Now, however, with the arrival of mini-computers and microprocessors, the trend is towards distributed computer capacity. Individual laboratories are acquiring their own mini-computers, while many pieces of equipment are now designed to operate with their own built-in microprocessors. (Most manufacturers of substantial pieces of equipment, for radioimmunoassay, for example, now provide floppy disks which allow the users to play games such as *Star Wars* as well as the more sober programs required to analyse the data, no doubt from compassion for those working in laboratories with time on their hands.)

This trend is probably inevitable, and unstoppable. For the user, the obvious advantage is to be fully independent, able to process one's data whenever necessary. Central machinery working on a batch-processing basis is necessarily by comparison an encumbrance. And there has grown up a certain air of disenchantment, at least among biologists and biochemists, with computer systems designed to be accessible by means of computer terminals in the laboratory.

Nevertheless, some of those concerned with the provision of computing facilities in general purpose laboratories are uneasy. Although the new small computers and microprocessors are cheap, the software required for specific specialized function may make serious demands of program-

ming skills. The end result, some gloomy calculations suggest, may be that the central cost to laboratories may be greater than at present. The alternative, that individual researchers should themselves become skilled at programming their own machines, is regarded with apprehension.

Laboratory administrators are also uneasy about the cost of measuring equipment with built-in microprocessors. The convenience and efficiency of these machines is beyond dispute. The suspicion is merely that the value added by the manufacturers of the equipment may not give the user a fair share in the cost benefits of the cheaper hardware now available.

The truth is probably that laboratory administrators are learning a painful lesson in economics. The development cost of software for a commercial measuring instrument is likely to be greater than that of writing a program in a particular laboratory. But the manufacturers of the measuring instruments with built-in microprocessors also find that demand is buoyant, suggesting that people able to afford the new equipment are prepared to pay for its convenience.

There are several well-attested illustrations of the costs that can be saved by home-made equipment. Thus one West Coast neurophysiologist, unable to afford from his current grant the \$20,000 to buy a suitable wave-form analyser, reckons that he was able to produce a better system for a quarter of the cost and at the same time produce a system of analysis better suited to his needs. The underlying dilemma is familiar — to what extent should research laboratories spend their time and energy on the construction of instruments which are nevertheless obtainable commercially.

Not all of the tasks of the general biological laboratory have as yet been delegated to the instrument manufacturers, however. Indeed, even in some of the most familiar tasks, the analysis of electrical output from physiological preparations for example, formidable intellectual difficulties keep cropping up.

The standard record in such experiments consists of a voltage (or several voltages) varying with time, on a time scale of milliseconds or thereabouts. Given a visual trace of such a voltage, it is possible with a ruler to estimate parameters such as the peak voltage, the time taken to reach that voltage and so on. Computer systems store the same information in digital form, and there are constant (if groundless) worries about the extent to which such data embodies "everything". More seriously, physiologists are likely to insist that some form of visual record should be obtained in parallel with whatever computerized analysis is found convenient. Machines may be able to analyse complicated sets of data once they have been provided with criteria for setting about this task, but

cannot recognize novel kinds of patterns.

So much is evident in the difficulties which keep cropping up in the research programmes intended to provide more sophisticated monitoring of the physiological condition of hospital patients, people in intensive care units for example. Many of the systems developed for this purpose depend on the recognition of abnormal trends in, for example, heart-beat rate, or variation from one cycle to another. One difficulty is that too little is known about the statistical properties of normal heart-beats to provide a sure yardstick for abnormality. The result is that false warnings abound, while few are confident that the systems now in use are capable of recognizing more than a small proportion of impending crises.

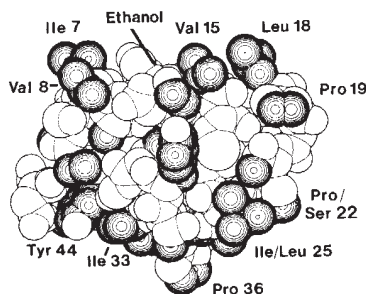
But what of the application of computers to tasks in biological research that have hitherto been beyond the scope of conventional techniques?

Many of the innovations in the past few years are comparable with those in other fields. It is unthinkable that X-ray diffraction data should, these days, be analysed by the traditional methods, and computer techniques are used as a matter of routine in construction of the models of most molecules, biological or otherwise. Quite apart from the way in which computing machinery has given people the courage to tackle problems that would otherwise have been neglected, the more economical use of data has made it possible to work with lower X-ray intensities.

Another development of this kind is in the construction of molecular models. Modelling biological molecules of any complexity is a formidable task, and this is the principal incentive for the development in the past decade or so of computer programs which can handle such tasks quickly but, more important, can in principle explore all possible configurations of a complicated molecule.

These are essentially data-handling tasks. The most recent development of this kind has been prompted by the sheer size of the nucleotide sequences now being accumulated. It is natural that these should be stored in computer form, and plans for the development of such a computer bank have been discussed between the National Institutes of Health in the United States and the European Molecular Biology Organization. One obvious elaboration of such a system is the use of computer programs for searching nucleotide sequences for particular configurations of nucleotides, for example for configurations that define the points of action of particular restriction enzymes. But the construction of a complete nucleotide sequence is not always straightforward, given that it may have to be inferred from the nucleotide sequences of overlapping fragments of a complete DNA molecule, which is why there are now ambitions to solve such essentially combinatorial problems by using computing machinery. □

Computer model of the protein crambin



Computers cheaper

One of the most surprising additions to the language brought about by the use of computers in high-energy physics is the verb "to Monte Carlo". (It is not clear whether the root of the verb should be hyphenated.) The verb is a reference to the now standard procedure by which somebody planning a high-energy physics experiment must recognize that he will not get the grant he needs unless he has first simulated the efficiency of his detector set-up by means of a computer calculation. The principle is that the path of any particle through the sequence of detectors and other recording devices that may be involved in an experiment will not only depend on the origin and momentum of the particle within a beam (whose dimensions are never negligible) but also on the strictly statistical chance that it will interact or decay at some stage. Analytical calculations are rarely possible; instead physicists use Monte Carlo calculations, in which the hypothetical paths of individual particles with parameters chosen by means of random number generators, and chances of interaction similarly determined, are followed through the equipment in the computer simulation. Successful applicants for grants and successful experimenters know that much depends on the skill with which their experiments have been "Monte Carloed" in advance.

This, however, is a relatively indirect use of computers in the high-energy physics laboratories. A much more direct use of computing machinery is in the control and management of experiments, and here perhaps the most striking application of computer techniques is in the control of the beam of 400 GeV protons in the SPS proton synchrotron at CERN.

The CERN system

The problem is easily stated. The machine consists of a vacuum chamber arranged roughly as a spherical ring 2,000 metres in diameter. The circulating beam of protons is intended to have a fixed orbit within the vacuum chamber while, to prevent collisions with the walls of the vacuum chamber and thus the loss of protons, the diameter of the beam itself should be of the order of 1 centimetre. But the beam circulates in the form of a pulse or package of protons, thereby increasing the electrostatic repulsion between them and adding to the inevitable space charge whose consequence is further to spread out the proton beam.

To avoid this difficulty, the proton synchrotron has been provided with a control system consisting of magnets for correcting the position of the beam within the vacuum chamber based on signals picked up by detectors elsewhere in the ring. It is tempting to describe this as a way in which the position of the beam is

adjusted in the light of information about its characteristics a few microseconds earlier, at the other end of the 2,000-metre diameter ring. In practice, of course, that description is an over-dramatization. It is entirely possible to accommodate departures of the proton beam from its intended position, or its spread beyond the intended width, for several orbits around the ring.

Computer control

Inevitably the beam adjustment has to be controlled by a computer. While in principle the system is not markedly different from those which make decisions on the ways in which detection equipment should operate based on observations of the properties of a particular particle entering the equipment, there is a sense in which the development of these systems has helped to make very high-energy accelerators feasible — or at least to reduce their cost or to improve their efficiency. The quality of a beam, after all, depends on its intensity (the number of particles in a single pulse) and on its small dimension. Too many particles lost in collision with the walls means too much beam intensity sacrificed. But the obvious solution to this problem, to use a larger vacuum chamber or stronger focusing magnets, adds

enormously to the cost of what is, after all, the most expensive part of such a machine, the magnet system in the poles of which the vacuum chamber sits.

Inevitably, such sophisticated control systems, already in use at CERN and at the Fermilab in the United States, will be incorporated in all large machines built in the years ahead. When funds are hard to come by, and when high-energy physicists must necessarily be seen by their sponsors to be making all possible economies in their proposals, it is clearly advisable to spend a little money on the computer control of the particle accelerator so as to save what may be a substantial percentage of the total cost of electromechanical construction.

Bubble chambers

These uses of computing machinery in high-energy physics are inconspicuous although now familiar. During the past 20 years, however, one of the most eagerly sought developments has been that of the computerized handling of the photographs of the tracks left by charged particles in bubble chambers — the modern versions of the Wilson cloud chamber. Tracks in bubble chambers consist literally of strings of bubbles forming in, say, liquid hydrogen at points where electrical ionization has been created. Especially with the improvement of accelerating machines and the inexorable trend towards higher energy, bubble chamber photographs (which come in stereoscopic pairs or

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

A sixteen-bit microprocessor — many thousands of transistors on a ¼ square inch of silicon

multiples) consist mostly of uninteresting tracks left by particles which have passed through the chamber unaffected (although perhaps deflected in one direction or the other, according to their charge, by the external magnetic field of a magnet). Traditionally, the analysis of such photographs, like the analysis of tracks in photographic emulsions, has been carried out in high-energy physics laboratories by small armies of part-time workers, usually women. Pressure to computerize the task of analysis has arisen not merely for economic but also for managerial reasons. A complete analysis of all the interactions recorded on a single bubble chamber image is too great to be attempted except in exceptional circumstances. The first attempts at simplifying the task required that the human operator should follow only tracks judged to be interesting with some device whose position in two directions could be recorded automatically, by pressing a button perhaps, stored in digital form and analysed to give a 3-dimensional reconstruction of the track.

In the past 15 years, however, the analysis of the complicated sets of images of which only some are interesting has been further simplified by the development of the so-called Hough-Powell device, essentially a means of scanning a plate or of scanning an image photometrically with a flying laser spot constrained on a predetermined raster. A digital output is thus provided in a simple form — each spot on the image is either black or white.

Limitations

In principle, appropriate analysis of two stereoscopic photographs of this kind can reconstruct the tracks of all charged particles passing through a bubble chamber. In practice, even with the use of computer hardware, the task is too slow and too demanding of storage space. Accordingly, it is still standard procedure for somebody, by means of visual inspection, to select those particular tracks in an image which are interesting enough to justify detailed analysis. One by-product of these computer systems is that they yield not merely a detailed geometrical reconstruction of a track but also a kinematical analysis of those particles involved. Whether they will ever be capable of operating entirely automatically, ignoring uninteresting tracks and singling out those which, by predefined criteria, are especially interesting or significant is another matter — and a problem in pattern recognition.

The development of these systems has been to an encouraging extent a collaborative business, in which the large high-energy accelerator laboratories have played a central part. In Europe, for example, one direct result of this collaboration has been the development of a set of standards (called CAMAC) intended to ensure compatibility of information processing equipment. □

Logic circuits before computers

When nobody was especially well supplied with equipment, but when nuclear physicists were better supplied than anybody — in the late 1940s and early 1950s — it was relatively easy to predict who the first users of computers would be. It is tempting to think, at least where the nuclear physicists are concerned, that they learned their habits during the Manhattan Project. In truth, the search for ways of applying computational techniques to the task of observing what you want to observe and not some tedious background signal came two decades earlier, in 1931.

Then, for practical purposes, there were merely a handful of techniques for observing the effects of energetic charged particles, one of which consisted simply of measuring the rate at which a charged electroscope would lose its charge. The most interesting technique was C.T.R. Wilson's cloud chamber, which had for the first time made it possible to visualize the tracks of charged particles travelling through its confined volume, but which to begin with could take snapshots of the flux of charged particles through it only at randomly chosen intervals. The chance of finding something interesting cannot have been much greater than that of finding a needle in a haystack.

Geiger counters, also widely used in the 1920s, were more discriminating than cloud chambers in that they would produce an electrical discharge only if the energy of the charged particle traversing them was greater than a certain amount determined by the quantity of precise composition of the gas with which the vacuum was filled — so much argon, so much nitrogen and so on. The snag with a Geiger counter (1930s vintage) is that it was simply a device for answering "yes" or "no" to the question "Where (and when) did a charged particle with more than the threshold energy pass by?"

But what more natural than to couple a layer of Geiger counters connected in parallel to the expansion mechanism of a cloud chamber so as to ensure that the cloud chamber photographs would include the track of at least one charged particle whose energy was greater than the threshold energy? This, in 1930, is what P.M.S. Blackett and G. Occhialini (then at the Cavendish Laboratory) set out to do. Blackett used ruefully to complain that if Occhialini had not taken a long holiday, they would have found the positron before P.W. Anderson.

From that point on, the use of elementary logic circuits, answering "yes" or "no" to questions such as "Has such and such a kind of charged particle passed through?", was inevitable. Arrays of Geiger counters working in coincidence ("A charged particle went in and another came out the other side") or

in anti-coincidence ("It went in, but nothing came out") were all the fashion. Chadwick proved the existence of the neutron by such a technique.

By the outbreak of the Second World War, people's ambitions had grown, while detection techniques had become more complicated — sophisticated is hardly the word. By encasing cloud chambers in magnetic fields, and arranging that the chamber would not be triggered into action unless a charged particle "went in" and emerged virtually unchanged in direction (indicating that its energy was very large), it was possible to ensure that all cloud chamber photographs included at least one track of an energetic particle (which is how the "strange" particles of matter were eventually discovered, in 1948).

Long before then, however, logic circuits and the development of detecting devices had made it possible not merely to



P.M.S. Blackett

record the arrival of a charged particle but to assign it to some range of energy. The outputs from the devices used for such purposes were essentially digital, so why should they not be accumulated in digital form? This is how it came about that nuclear physics laboratories in the 1950s, but especially those connected with semi-secret enterprises, were the first to be blessed not merely with huge racks of electronics fitted out with tube-driven logic circuits but with slave electric typewriters as well, recording the data the detectors generated. (Most secretaries in those same laboratories still used manual machines.)

Frivolity apart, the outcome has been important. The high-energy physicists have been among the leaders in the use of computers for the control of experimental equipment. Most of the detection equipment now run in connection with high-energy particle accelerators consists of an elaboration of the counter-controlled cloud chamber now exactly half a century old. The difference is that the primitive logic circuits of the old devices have been replaced by computing machinery. □

Astronomy by numbers

Computers have taken astronomy by storm in the past decade or so, and for obvious reasons. Need has been matched by opportunity, but astronomers have also been temperamentally attuned to computerization by the space programmes of recent years. Once it became necessary to work with data transmitted in digital form by telemetry from rockets and satellites, it was certain sooner or later to become habitual to use the same techniques for dealing with other sources of data.

Astronomy has traditionally been profligate in its regard for data. A single plate from a large Schmidt camera, covering a patch of sky just six degrees in each dimension, may include up to 10^6 images of different kinds. Merely recording the positions of these images would occupy a skilled technician for the best part of a lifetime. The libraries of the world's observatories most probably contain enough information to keep the world's astronomers usefully employed for the rest of their lives — perhaps on problems of no particular interest.

The simple solution to the problem of counting stars and measuring plates is similar to that to which high-energy physicists were forced in the 1960s, when it became impracticable to measure the tracks of fast particles in bubble chambers

by hand and eye. The first need is a device for converting star images into numerical records of the density of exposure at different points on a plate. Plate reading systems have been developed at several observatories. As always, the tricks lie not in the hardware but the software — in this case the rudimentary programs for telling which image is what. Is it a star? Or a galaxy (and if so which kind)? Or something else?

One neat way of describing the benefits of these systems is that it has taken some fifty man-years of programming time to develop the software used with the COSMOS plate measuring system at the Royal Observatory, Edinburgh and that the result is to make possible the measurement of the images on a single Schmidt plate in one day, not one lifetime (also fifty man-years). This figure of merit is also an index of the true cost of these devices — £500,000 is probably an underestimate of the true cost of developing COSMOS.

The other component in the recent computerization of astronomy is the digitization of the data base. The most striking illustration so far of the benefits of this technique is the Einstein X-ray satellite (now, happily and surprisingly, restored to working order). The essence of the X-ray telescope is a system of conically shaped

glass surfaces from which X-ray photons are deflected at more or less grazing incidence. The detectors are inherently electronic, borrowed from nuclear physics. So the data are inherently digital in form, literally consisting of the spatial coordinates in the image plane of individual X-ray photons. So why not record not merely the angle from which these photons have arrived but the time of their arrival and even information about the energy of the incident photons?

In practice, the Einstein satellite is equipped with a variety of X-ray detectors which separately provide different kinds of data. The detector with the highest resolution, called the High Resolution Imager, is designed to record the arrival of X-ray photons on the MgF_2 plates from which cascades of electrons are generated to within 1 second of arc and within 8 milliseconds. The working spatial resolution of the instrument is less than this (some tens of seconds of arc) because of the properties of the mirror system, but this high resolution instrument provides no spectral information worth speaking of, with the result that the computer systems store merely spatial and temporal information.

Other instruments in the Einstein Observatory, which can generate spectral information (but whose spatial resolution is not as high), yield data streams from which spatial, temporal and spectral information can be regenerated. The

Some things old, some new

Electronic computers are merely the latest stage in the computerization of the laboratory. The earliest aids to laboratory computation were probably the astrolabes and armillary circles going back to the antiquity of astronomy. The first explicit

aids to laboratory computation were probably Napier's logarithmic systems of coloured rods (called "bones") introduced early in the seventeenth century, and quickly followed by Pascal's logarithmic scale, much used (with the help of a pair of dividers) by navigators in the eighteenth century.

The needs of astronomers stimulated interest in what may have been the first computerized aid to laboratory observation — sky-following telescopes. Robert Hooke appears to have been the first to have conceived of this innovation, but his proposal for a clockwork-driven telescope was built for G.D. Cassini at the Paris Observatory and commissioned in 1682. This was a huge device, with the objective mounted at the centre of a clockwork-driven structure and with the

eyepiece mounted separately at a distance of more than 100 feet. Cassini's telescope was known as his *machine parallatique*.

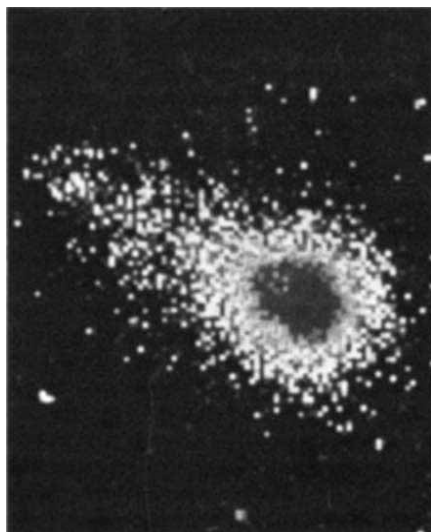
Attempts at the further development of clockwork-driven telescopes appear to have been few and far between in the eighteenth century. Such instruments seem to have come into their own only in the early decades of the nineteenth century.

Dr John Darius (Science Museum, London) writes:

Driving telescopes around their polar axis to follow the diurnal apparent motion of the stars is useful for visual observations but absolutely vital for astronomical photography. The earliest drives, dating from the beginning of the nineteenth century, were purely mechanical, depending on a clockwork mechanism driven by weights to rotate the telescope in the equatorial plane at the rate of 15 degrees of arc every hour. By around 1870, telescope drives regulated by an electric pendulum clock had been invented, but by 1910 electric motors were coming into regular use. Nowadays digitally controlled synchronous motors as uniform as the frequency of the mains supply have almost wholly replaced their historic forebears. The most modern systems now use servomechanisms which are controlled by computer. □



W.H. Smyth's clockwork-driven telescope, 1829



An image of comet Bradfield taken by the Fine Image Sensor of the IUE satellite observatory

practical value of this complication is that some X-ray sources are especially interesting because of their rapid fluctuation with time. Thus some X-ray sources that consist of binary star systems in which material is being transferred from one star to the other intermittently fluctuate in X-ray intensity on time-scales of a few minutes. Others, for example quasars, also fluctuate in time. Some of the earliest observations with the Einstein Observatory (see, for example, Tananbaum, H. *et al.*, *Astrophys. J. Lett.* **234**, L9; 1979) were used to demonstrate variability of X-ray intensity from distant quasars over a period of a few hours on the basis of a mere 50 or so X-ray photons.

It goes without saying that the planning of a series of observations with such a satellite requires that the on-board computers should be explicitly instructed on what information to store, and in which form, while the planning of the way in which a high-resolution instrument scans a small patch of sky requires some forethought about the way in which the data will be analysed. Programs by which the field of view moves in overlapping steps across the sky are often advantageous. It is commonplace that observations of particular objects are pre-programmed not for arbitrary periods of time but in such a way as to yield statistically significant results (which entails concurrent measurement of the background radiation).

Astronomy as a whole now seems destined to move in these directions, helped by the development of a variety of detection devices whose chief purpose has been to improve the sensitivity of telescopes (optical, radio, infrared, what-have-you) but which as a by-product yield a stream of numbers as the crude data. Quite what balance will in the end be struck between positional, spectral and temporal information is, however, still very much a matter for astronomers to decide. For the plain truth is that while it may in principle

be possible to record every detail of every quantum event occurring at the light-sensitive surfaces of the image intensifiers, there is no hope of being able to store such a wealth of data.

Accordingly, there is no sense in which the computerization of astronomy can be said to have robbed the craft of the need for discrimination and judgement. Although the coming of more efficient light detectors has often made possible dramatic reductions in the length of time needed to gather significant information about some distant object, observing time is still at a premium. The need for deciding clearly in advance what questions it is hoped to answer, and for collecting the smallest amount of data required, has therefore not diminished. Indeed, it has been intensified by the ease of discrimination in the collection of data.

The bugbear that remains is that the craft has become almost inextricably bound up with computer programming. Research students have repeatedly found themselves caught up in projects for writing programs to recognize, say, elliptical galaxies, or to measure the ellipticity of globular clusters. Educative though these exploits may be, there is obviously a danger that too much creativity may be lavished on what is essentially repetitive work.

This is part of the incentive for the development in the United Kingdom of a linked network of six astronomy centres in a project funded by the Science Research Council under the trekky name of STARLINK. The plan is that five British centres (the Royal Greenwich Observatory, the Royal Observatory Edinburgh, University College London and the Universities of Manchester and Cambridge) should be provided with identical computer systems and linked by low-capacity data links to a coordinating centre at the council's Rutherford-Appleton Laboratory.

The objective is to make possible the more economical use of people's pro-

gramming energies. Schemes for the development of novel programs will be worked out on an *ad hoc* basis by the collaborating laboratories. Programs but also images (containing 500×500 picture elements or pixels to begin with, perhaps 16 times as many later on) can be transmitted from one centre to another. There are some ambitions that the system may be extended internationally, to Europe, possibly the United States and, with luck, to Australia.

The STARLINK system is interactive in that the DEC VAX computer system provides astronomers with a graphic display of an image which can be manipulated this way or that, to yield profiles across an object for example. Much of the work so far, however, has been the development of programs for the analysis of spectra. How does one extract from a record of intensity as a function of wavelength numerical data for line frequencies and line depths? Superficially, the problem resembles that which crops up in quite different fields. In medical physics, people are forever trying to decide how best to characterize heartbeats.

The system is not yet fully a going concern, although it is hoped that a full package of software will be available "in months rather than years". For some purposes, however, the response of the system is plainly not fast enough — people who want to follow the motion of features in the atmosphere of Jupiter, for example, need more sophisticated image processing, with a much faster response. This is why, for instance, some of those concerned with the processing of data from the recent series of planetary probes have independently developed more sophisticated image processing systems. One such is the IPSIS system at University College London, which gives scientists access to a large number of satellite images stored in digital form. The interactive nature of the system permits a detailed examination of all aspects of the data. □

Algebra made mechanical

The notion that computers are mere number crunchers is fast being exorcised by one small group of people — the computer algebraists. In the past twenty years, several classes of problems which require analytical and not numerical solutions have been tackled successfully. Now, for example, there are computer programs that can yield analytic expressions for indefinite integrals a good deal more reliably than the "random undergraduate", as one practitioner puts it.

Neither the need nor the feasibility of these tasks is self-evident. And, as things are, the community of computer algebraists is small and scattered. Yet, in some fields, it is now commonplace for people with time-consuming algebraic calculations on their hands to use a computer program for the purpose. And

the fields in which these techniques are being used are so diverse that there seems every prospect that computer algebra will spread.

The feasibility of computer algebra is demonstrated easily enough — perhaps most easily by the manipulation of polynomial expressions such as $a + bx + cx^2$. Adding second-degree polynomials together requires an algorithm for adding together the coefficients. Multiplication and subtraction are similarly straightforward. Division is (as always) more difficult, but there is no reason why the rules of school algebra should not again be applicable.

Producing analytical forms of indefinite integrals, or analytic solutions to differential equations, is necessarily more difficult, not merely because the computer

programs used for these purposes must in some sense "know" the essential mathematical properties of functions such as $\sin x$ or e^x , but also because the analogy with the way in which students are taught to solve such problems (by classification around solvable standard forms) is not sufficiently general.

Yet programs for these purposes have been developed at several centres. Put to work on sets of problems taken from standard textbooks, they typically yield success rates of 97 or 98 per cent — they have more than the competence expected of undergraduates but are not yet good enough to satisfy the purists. But these, of course, are not the fields in which computer algebra is likely to make its mark.

There are many more telling applications of the technique, in fields where in the past heroic individuals have battled with forbidding mountains of algebra or have, as likely as not, decided that discretion is the better part of valour.

The successes

One of the first and most striking successes in this field was in celestial mechanics — the theory of the Moon. The problem is straightforward but daunting. Since Newton's time, a knowledge of the geocentric longitude of the Moon as a function of time has been an essential ingredient of the craft of navigation. To state the problem is simple enough — it is a classical problem in Newtonian mechanics. It is, however, a three-body problem even without the gravitational effects of Jupiter and the other planets and is thus incapable of yielding an exact solution.

In the circumstances, only perturbation methods of solution will suffice. Numerical solutions, widely used, have one serious and for some purposes insuperable drawback — errors accumulate horrendously as the span of the time variable increases, so that it is not, for example, possible to be sure of where and when some past eclipse should have occurred. Accordingly, there has always been great interest in analytical solutions of the problem of the Moon — essentially approximations in the form of convergent series whose terms are individually analytical.

The only successful attack on the problem "by hand", so to speak, was that of Delaunay, whose two huge volumes on the subject were published by the French Academy in 1860 and 1867. The first hundred pages of the first volume are entirely given over to a representation of Delaunay's analytical expression of his "disturbing function" (essentially the departure of the gravitational potential of the Moon from simple Newtonian calculations of the potential due to the Earth). This seems to have represented literally a lifetime's algebra — Delaunay carried through his calculations to the seventh order in small quantities. Now, the same

calculation can be run on a computer in a few tens of seconds, and to a higher order of approximation. Delaunay's descendants need not be ashamed that the first computer programs showed that his algebra had not been immaculate.

Similar analytical solutions of more intricate (and previously intractable) problems in celestial mechanics are now matters of routine — for example, the movement of putative Earth satellites about the Lagrangian points at which the gravitational forces of the Earth, Sun and Moon cancel out. The practical importance of these developments is considerable, if only because the error of an analytical approximation is more easily assessed than that of a numerical approximation.

General relativity

Curiously, as it may seem, one of the other well tried (and much used) applications of computer algebra is in general relativity the problem is this. Einstein's general theory of relativity is general in the sense that it is possible to derive the gravitational field equations, typified, for example, by the gravitational forces on a test particle itself of negligible mass from an analytical expression for the metric of space-time. The simplest possible form of the metric is

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2 t^2$$

where c is the velocity of light, x , y and z are cartesian coordinates and t is time. Substituting this metric into Einstein's general theory merely recaptures the kinematic aspects of special relativity, as of course it should.

More generally, the metric of space-time is written

$$ds^2 = G_{\mu\nu} dx^\mu dx^\nu$$

where the indices μ and ν run from 1 to 4, x^μ is one of the four coordinates of space-time and the quantities $G_{\mu\nu}$ are, in general, functions of these coordinates. The underlying convention adopted by the relativists is that indices which occur twice represent summations — in other words, the general metric consists of 16 terms of which the first is

$$G_{11} dx^1 dx^1 = G_{11} (dx^1)^2$$

Regarded as a general theory, the General Theory is intractable. The field equations, the differential equations which predict how particles should behave, are nonlinear and therefore soluble only with special assumptions. So the standard way of tackling problems in general relativity consists either of guessing that a metric might illustrate some particular property of a gravitating universe, and to calculate from that the field equations, or alternatively to look for approximate solutions of the field equations, valid in some narrow regions of space and time. In either case, the hapless relativist finds it necessary to calculate what are called the Christoffel symbols of which in general there are 64, each of them a combination of three first partial derivatives of the functions which define the metric. That, however, is only a

beginning. The next step is to calculate the Riemann-Christoffel tensor, which is in general a tensor with 256 distinct components, each of which is a linear combination of partial differential coefficients of the Christoffel symbols.

The sheer algebraic complexity (and tedium) of these calculations explains why only a few relatively simple metrics have been used in general relativity. Even the simplest of the more general metrics, the general case of an orthogonal metric

$$ds^2 = D(dx^1)^2 - A(dx^1)^2 - B(dx^2)^2 - C(dx^3)^2$$

in which the symbols D , A , B and C stand for functions of the four coordinates x^i etc. presents formidable problems of computation. Herbert Dingle in 1933 published a list of the Christoffel symbols (*Proc. natn. Acad. Sci. U.S.A.* **19**, 559; 1933) which represented the fruit of several years work and which remains a standard reference list. The likelihood that such calculations would now be attempted with more general metrics is exceedingly small, which is why computer programs for making these calculations are widely used.

Probably the most widely used set of computer programs with applications in theoretical physics is, however, that called SCHOONSCHIP, developed in the late 1960s at CERN in Geneva as a means of calculating physical quantities in quantum electrodynamics. Briefly, such problems invariably create the need to calculate the trace (the sum of the diagonal elements) of a four-dimensional matrix which is itself the sum of what may be a large number of four-dimensional matrices obtained by multiplying together the so-called Dirac matrices. Although the elements of the Dirac matrices are simple (either 0 or 1), the algebraic task of computing by hand the product of say eight of them is enormous, and the chance of making a mistake is very great. (In practice, the multiplication of Dirac matrices goes hand in hand with the evaluation of four-dimensional integrals, now also computerized.) In some applications, as for example in the application of similar techniques to unified field theories or quantum gravity, many vital calculations have not been carried out simply because of the complexity of the task. Inevitably, with the passage of time, programs such as SCHOONSCHIP have therefore become widely used.

Versatility

These three applications in theoretical physics of computer programs are tangible and valuable. But if the object of computerizing algebra is to provide theoretical physicists with what Professor Joel Moses has called an algebraic assistant, surely computer systems should be more versatile. This consideration underlies the attempts which have been made since the late 1960s to develop general-purpose algebraic computer programs, the best known of which is called MACSYMA

and lives in a PDP-10 computer at the Massachusetts Institute of Technology (MIT). A similar computer system called REDUCE was developed during the same period at the University of Utah.

In principle, these programs are designed to deal with whatever algebraic problems may be thrown at them. In their development, handling polynomial algebra and matrix multiplication was relatively straightforward. The problem of integration, requiring that the program should make decisions about the most appropriate way of solving a problem, was necessarily more difficult. Even differentiation creates problems when irrational functions are involved. Now, however, MACSYMA has more than a thousand users, many of whom make use of it by means of international communications systems such as the ARPA network. Versions of the program are soon to be made available in a dialect of the computer language LISP, when MACSYMA is likely to be as easily available as its competitor, REDUCE.

Data stores

Inevitably, with such ambitious programs, the storage capacity required is very large. For practical purposes, the program has to be provided with a comprehensive knowledge of mathematics. Professor Moses estimates that MACSYMA at present uses a million 32-bit words of storage capacity but users may be able to get by with less if they make do without some of the more general functions.

Storage capacity appears to be a recurring problem in computer algebra, not merely for comprehensive systems such as MACSYMA and REDUCE, but also for more straightforward algebraic programs. The reason is easily understood. If the objective, for example, is to compute the polynomial expression $(a + bx)^n$, it is clear that the result must be a polynomial of degree n with $n + 1$ distinct coefficients. But the number of terms that arise by multiplying together one of the two elements in each of the n factors is 2^n . If n were anything like as large as, say, 100, it would be physically impossible to store the terms separately in the computer. In practice, of course, this particular computation is straightforward. But the problem of storage does compel the designers of algebraic programs constantly to be on the lookout for algorithms that make economies. For the simple binomial expansion above, for example, it plainly makes sense that the factors should be multiplied together consecutively. Nevertheless, program designers have found it necessary to provide overspill storage in case a problem should have need of it, often by means of buffer storage on magnetic tape which has then to be searched regularly and reincorporated in the accessible memory of the computer if the program so decides.

Users of both comprehensive programs such as MACSYMA and more specialized

Plenty of programs

Altogether there are probably more than 50 computer algebra systems now available. The following are the best known or most widely used.

- **MACSYMA.** Developed since 1968 at the Mathlab at the Massachusetts Institute of Technology by Professor Joel Moses and his colleagues. The objective has been a comprehensive system, capable of solving all algebraic problems. Development continues. Hitherto the system has been accessible only at MIT but the program is soon to be made available to outside purchasers. It is slower in specialized applications than purpose-designed computer programs.

- **REDUCE.** Largely developed by Professor A.C. Hearn, originally at Stanford University and then at the University of Utah. (Hearn is now at the Rand Corporation, Santa Monica.) The objective is a comprehensive system. Development has been dependent on collaboration and, collaborators say, has profited therefrom. The program is usable on a variety of machines (although some would-be users have found that adaptation is difficult). The system is still being defined and made more comprehensive. Users say that newly available programs are much more sophisticated than earlier versions.

- **SCHOONSCHIP.** Developed by Veltman at CERN in 1972 with the objective of solving problems in quantum

electrodynamics. Within its range of applications the system is fast and its demands on computer storage modest. The program is easily usable on large CDC computers but less easily on machines of smaller word length.

- **CAMAL.** The system developed at the University of Cambridge in the 1960s for solving problems in celestial mechanics. Fast for what it does and modest in demands on storage. Still widely used but now losing ground to more sophisticated systems, for example that developed by the Boeing Space Sciences Division.

- **MUMATH.** Developed by Stoutemyer at the University of Hawaii and intended for use with microprocessors. Suggested uses include teaching but also an application to small research problems, and in particular for system design.

Most computer manufacturers have algebraic programs as do the Bell Telephone Laboratories.

Such coordination as there is of academic interest in algebraic computing is undertaken in the United States by SIGSAM, a special interest group of the Association for Computing Machinery, which organizes meetings (the next at Snowbird, Utah in August this year) and which also publishes a quarterly bulletin. In Europe, coordination is through the Symbolic Mathematics Committee, organized by J.A. van Hulzen (T. Hogeschool Twente, Enschede, The Netherlands). □

systems such as SCHOONSCHIP testify to the complementarity of the two kinds of systems. The comprehensive systems are inevitably slower than the specialized systems and also less convenient to use in practice. On the other hand, access to the comprehensive systems can be of great value in the amendment of an existing program. Users also say that there are pitfalls to be avoided when programs such as REDUCE are purchased on magnetic tape and used with computing machinery different from that on which they were originally developed, which is part of MIT's justification for not hitherto making MACSYMA widely available.

Curiously enough, although the algebraic computer programs are in general demanding of computer storage, they are relatively undemanding of programming time. MACSYMA, largely the product of people working at MIT since 1968, is estimated to have taken between 50 and 100 man-years of programming effort. Other systems designed for specialized purposes can be relatively simple to design and write. One practitioner, for example, says that the time required to write a program for simple polynomial algebra is not much more than a quarter of an hour. No doubt the relatively undemanding nature of some of the programming required reflects the character of the problems to which

computer algebra has so far been applied — their intrinsic tedium. And, of course, nobody pretends that writing programs to carry out integration, to solve differential equations or to tackle problems in mathematics such as the enumeration of finite groups can be solved in a straightforward way.

What the future holds for the applications of computer algebra remains to be seen. Already, however, some of the practitioners hold that the availability of these techniques provides a means by which shortages of mathematics teachers, even at the school level, might be remedied. Indeed, the University of Hawaii has developed a computer algebra program (commercially available) that can be run on home computers, while there are ambitions to use MACSYMA in the education of undergraduates.

What educationists will have to say about this prospect is unclear. The hand calculator, once considered a threat by elementary school teachers, seems now to have been recognized as a stimulator of interest. It is harder to see how more formal computer programs will be used in teaching algebra, and it is widely thought that too indiscriminate use of computers for algebraic computations in research may blind users to the physical significance of what they are doing. □

REVIEW ARTICLE

Can dietary beta-carotene materially reduce human cancer rates?

R. Peto*, R. Doll†, J. D. Buckley‡ & M. B. Sporn§

* Imperial Cancer Research Fund Cancer Studies Unit, Nuffield Department of Clinical Medicine, Radcliffe Infirmary, Oxford OX2 6HE, UK

† Imperial Cancer Research Fund Cancer Epidemiology Unit, 9 Keble Road, Oxford OX1 3QG, UK

‡ Department of Biomathematics, University of Oxford, Oxford OX1 2JZ, UK

§ Laboratory of Chemoprevention, National Cancer Institute, Bethesda, Maryland 20205, USA

Human cancer risks are inversely correlated with (a) blood retinol and (b) dietary β -carotene. Although retinol in the blood might well be truly protective, this would be of little immediate value without discovery of the important external determinants of blood retinol which (in developed countries) do not include dietary retinol or β -carotene. If dietary β -carotene is truly protective—which could be tested by controlled trials—there are a number of theoretical mechanisms whereby it might act, some of which do not directly involve its ‘provitamin A’ activity.

THE possibility of discovering anticancer substances that can be prescribed rather than carcinogens that must be proscribed is attractive, for people may be more willing to accept prescription than proscription. If inhibitors of the late ‘stages’ of neoplastic transformation could be found, they would have the added advantage of conferring their benefits without an impracticably long latent period. Since there is growing evidence that retinoids and perhaps also carotenoids can be anti-carcinogenic, we review the intake and metabolism of ‘preformed vitamin A’ (retinal and retinol and its esters) and of the carotenoids (some of which exhibit ‘provitamin A’ activity), and we review the epidemiological evidence correlating blood levels or dietary intake of one or other of these with low cancer risks. Two independent questions will be considered.

First, within the range of variation of total blood retinol levels consistent with the avoidance of hyper- and hypo-vitaminosis A, is there any material variation in human cancer risks attributable to a protective effect either of unbound retinol or of carrier-bound retinol in the blood, and, if so, what are the important extrinsic determinants of the concentration of carrier-bound retinol in human blood? (Paradoxically, within the wide range of dietary intake rates consistent with avoidance of both hyper- and hypo-vitaminosis A, the amount of retinol derived directly or indirectly from the diet makes no material difference to the blood levels of carrier-bound retinol.)

Second, within the range of blood levels of β -carotene (or some other carotenoids) which are insufficient to cause cosmetically unacceptable skin yellowing, is there any material variation in human cancer risks attributable to a protective effect of circulating carotenoids, and, if so, what are the important extrinsic determinants of the concentration of β -carotene and other carotenoids in human blood? (In contrast with blood retinol, blood carotenoid levels are directly determined by dietary carotenoid levels, although they may also be affected by certain chronic disease states.)

The evidence relating to the above two questions will be reviewed under three headings:

(1) The qualitative difference between the pharmacodynamics of β -carotene and of retinol¹⁻³ in the human body: which

suggests that the possible relevance of these two components of ‘dietary vitamin A’ to cancer prevention should be assessed separately.

(2) The inhibition by retinol and various synthetic retinoids of the later, rather than the earlier, processes of carcinogenesis in experimental studies with animals and cultured cells: which suggests that if such agents have any substantial anticancer effects on humans, a reduction in risk should be measurable within about 5 years, instead of after a latent period of some decades, as might be expected for inhibitors of the ‘early’ processes of carcinogenesis.

(3) The tendency, observed in many epidemiological studies, for some human cancer risks to be inversely associated with blood retinol, and with blood or dietary β -carotene.

Pharmacodynamics

‘Vitamin A’ is a nonspecific term embracing two different families of dietary factors, one family comprising the various types of ‘preformed vitamin A’ (chiefly retinyl esters, but also retinol and retinal), and the other family comprising the various types of ‘provitamin A’ (chiefly β -carotene, but also those other carotenoids that can be metabolic precursors of retinol). If retinal (the aldehyde $C_{20}H_{28}O$) is written $X=O$, then retinol, the corresponding alcohol, is $H-X-OH$, and β -carotene is the polyunsaturated hydrocarbon $X=X$, which may be oxidized by the enzyme dioxigenase to two molecules of retinal, chiefly in the intestine but perhaps also in the liver⁴.

The various types of preformed vitamin A are converted efficiently by the human body into retinol, and a variable fraction of dietary β -carotene is converted in the human intestine into retinal. Nutritionists estimate² that even among poorly nourished people β -carotene usually yields only about one-sixth of its weight of retinol; furthermore, the fraction converted decreases as the amount of β -carotene in the diet increases so that even massive doses of β -carotene will not yield enough intestine-derived retinol to cause systemic toxicity (‘hypervitaminosis A’), unless massive amounts of preformed vitamin A are also taken. Despite this inefficient intestinal conversion of β -carotene, and the even less efficient intestinal

Caution

Unwary readers (if such there are) should not take the accompanying article as a sign that the consumption of large quantities of carrots (or other major dietary sources of β -carotene) is necessarily protective against cancer, and the correlation between blood retinol and cancer avoidance is, for the time being, *sub judice*.—Editor, *Nature*.

Table 1 Prospective studies of cancer in relation to total blood retinol in stored blood samples taken from apparently healthy people

Country	Relative risks (in sequence from low to high blood retinol quintile)					No. of cancers (all sites)
	5.3:	2.9:	3.8:	3.0:	1*	
Georgia, USA ³¹	5.3:	2.9:	3.8:	3.0:	1*	85
United Kingdom ³²	2.2:	1.7:	1.6:	1.1:	1	86

* Top quintile based on only seven cancers, and so unreliable.

conversion of most other carotenoids, many people in rich countries and most people in poor countries derive the majority of their retinol from the provitamin A carotenoids in plants. However, some of the dietary β -carotene and other carotenoids is absorbed unchanged from the intestine, circulates around the body, and, being fat-soluble, is stored in adipose tissue. Various factors then affect its metabolism and excretion (for example, diabetes⁵, thyroid disease⁵ and, paradoxically, anorexia nervosa^{6,7} will elevate blood β -carotene levels), but other things being equal blood and adipose tissue levels of β -carotene and other carotenoids are directly related to the amounts ingested over the preceding few weeks⁸. Normally, a daily average of only a few milligrammes of β -carotene is consumed. This makes human fat appear yellow but is not sufficient to cause noticeable yellowing of the skin. However, if a few dozen milligrammes of β -carotene per day are consumed for several weeks by Caucasians, 'hypercarotenaemia' may develop, that is, skin yellowing may become cosmetically noticeable. No medically significant side effects of hypercarotenaemia are known⁹, and the condition disappears within a few weeks of reversion to a normal diet.

The mechanisms for storing, transporting and excreting retinol are more specific than those thus far discovered for β -carotene. Intestine-derived retinol enters the blood as 'unbound' retinol, esterified but not bound to the specific carrier protein for retinol (see below). Unbound retinol is rapidly absorbed from the blood by the liver, where it is stored. To transport controlled amounts of retinol from the liver to the peripheral tissues where it is needed, the liver synthesizes a carrier protein called retinol binding protein (RBP) at a rate which, unless the liver's stores of retinol are almost completely depleted, is not strongly dependent on the amount of retinol currently stored in the liver. (The total rate of loss of retinol from the liver is, however, directly dependent on the total amount stored in the liver^{10,11}, so other, presumably degradative, pathways must also exist, and will provide some homeostatic control against unlimited accumulation of retinol in the liver.)

RBP molecules can leave the liver as soon as they have picked up a retinol molecule from the liver stores; but if the liver stores are severely depleted RBP builds up in the liver, waiting for some unbound retinol to be derived from the intestine. This RBP build-up is characteristic of hypovitaminosis A, that is, vitamin A deficiency. At the opposite extreme, if due to regular heavy intake of preformed vitamin A the rate of derivation of unbound retinol from the intestine consistently exceeds the maximum possible rate at which the liver can get rid of retinol (both by steady carrier-bound release and by concentration-dependent degradation), the liver's retinol storage capacity will eventually be exceeded, damaging the liver and allowing significant amounts of unbound retinol (not bound to RBP) to circulate. This may lead to serious symptoms of hypervitaminosis A, since unbound retinol should never circulate in significant quantities except in brief transit from the intestine to the liver. However, most of the population of developed countries has a diet-derived intake of retinol within the wide range which avoids the well-defined physiological conditions of hypovitaminosis A (unbound RBP in the liver) and hypervitaminosis A (unbound retinol in the blood). Within this wide range, blood levels of carrier-bound retinol (retinol bound to RBP) are not materially dependent on the amount of free retinol derived from the intestine. For example, among Swedish acne patients an

increase of the normal dietary dose of retinol by a few thousand per cent for a few weeks increased carrier-bound retinol by only about 10% (ref. 12). Likewise, in New York, liver retinol levels, which are directly related to dietary intake rates, are uncorrelated with blood retinol levels¹³.

However, blood levels of carrier-bound retinol do have extrinsic determinants. For example, oral contraceptive users have substantially elevated blood retinol levels^{14,15}, and there are presumably also various other hormonal or non-hormonal determinants of the consistent differences between people in their long-term average carrier-bound retinol levels (including¹⁶ diseases of the liver, thyroid and kidney), although none seem to be well understood. The key finding for this whole review is that for most people in developed countries the amount of intestine-derived retinol is not one of the important determinants of the blood levels of carrier-bound retinol (or of total retinol, since free retinol is usually present in the blood only at much lower concentrations than carrier-bound retinol is).

Possible short latency of antitumour effects of retinoids and related compounds: The ICRDB¹⁷ have compiled summaries of some 324 cancer research papers concerned with retinol, other retinoids, β -carotene or other carotenoids, and the October 1979 issue of *Federation Proceedings* contains several further relevant papers. Together, these show that various retinoids can reversibly suppress the malignant behaviour of cultured cells that have been transformed by viruses, chemicals or ionizing radiation; they can delay or prevent the onset of cancer in experimental animals that have been treated previously with DNA-binding carcinogens; they can cause human skin keratoses or oral leukoplakias to regress; they can delay the appearance of transplanted tumours; they can prevent 'tumour promoters' from eliciting papillomas on mouse skin; and they can inhibit some of the very specific biochemical effects of tumour promoters on cells *in vivo* or in culture. Although inappropriate use of retinoids may do more harm than good¹⁸, it thus seems reasonable to hope that exposure of human tissues to retinoid-like activity can be manipulated to reduce cancer risks¹⁹.

Neoplastic transformation can often be subdivided into 'stages', the early stage(s) involving only partial alteration of a normal cell and the later stage(s) involving conversion of a partially altered cell into a fully neoplastic cell. Finally, to be effective the neoplastic cell must not be eliminated, and must proliferate into a pathological tumour. Early- and late-stage changes are both necessary (so halving either might be an equally effective way of halving cancer risks), but typically have different causes^{20,21}. The demonstrated protective effects of retinol (and, perhaps, β -carotene) are such as to suggest that it is the later stages of neoplastic progression which are chiefly affected, together perhaps with the final process of tumour growth. There might often be a latent period of a few decades before a doubling or halving of the early processes had any material effect on human cancer risks, but, to judge by the rapidity with which the effects of stopping smoking are evident, a doubling or halving of the late processes might have noticeable effects on human cancer risks within about 5 years, while effects on the final process (proliferation) might be even more rapidly evident. So, human cancer risks might be noticeably influenced by the average levels of blood retinol within a period of as little as 5 or 10 years. This conclusion greatly facilitates observational epidemiology and may even make practicable a randomized human evaluation of the anticancer effects of such agents (or of any other agents which inhibit either the late or the final processes).

By contrast with the retinoids, the question of whether dietary β -carotene or some other carotenoids can affect cancer risks has not received nearly as much experimental attention. Possible mechanisms to consider include (1) a direct retinoid-like effect of some carotenoid(s) on cellular differentiation in the target tissues, (2) conversion in the target tissue of some carotenoid(s) into molecules with some such retinoid-like effects, or (3) protection by carotenoids of the target tissues via mechanisms not

related to control of cellular differentiation (for example, by enhancing some immunological function^{22,23}, or by quenching singlet oxygen, a highly reactive excited molecular species which may be generated as a toxic by-product of many normal metabolic processes in both animals and plants: indeed, β -carotene is the most efficient quencher of singlet oxygen thus far discovered²⁴, and in plants the chief function of carotenoids may be to quench the singlet oxygen produced by photosynthesis²⁵). Before possible mechanisms are discussed, however, it must be asked whether any carotenoids have any general anticancer effects whatever.

We know of only nine experiments on the modifying effects of systemic β -carotene²⁶⁻²⁸ or of carrot extracts²⁹ on induced tumours. In each experiment a protective effect was claimed, but lack of published details of the exact experimental circumstances makes some of these claims difficult to evaluate. If the reported²⁸ protective effects of dietary β -carotene against tumour induction by transplanted cells or viruses are confirmed, this may merely reflect the reported^{22,23} enhancement of certain immune parameters by dietary β -carotene, the relevance of which to human cancer remains uncertain.

If the reported protective effects of systemic β -carotene against chemically induced skin cancers^{27,29} are confirmed (and are not trivially attributable to physical changes in the skin) this would be of more interest than protection by systemic β -carotene against skin cancer induction by UV light^{26,27}, for the latter may at least in part reflect a photoprotective rather than an anticancer effect. Further studies in all these areas will be needed before any general protective effects of β -carotene in experimental animals can be accepted. (No excess of tumours was reported in the chronic toxicity test of β -carotene³⁰, but details were not provided.)

Epidemiology

The foregoing laboratory evidence has been introduced not because it proves that β -carotene will have an anticancer effect, but chiefly as background material to aid interpretation of the epidemiological observation that people with above average blood retinol levels or above average β -carotene intake rates are at below average risk of cancer. Although the available epidemiological data (see below) indicate that both factors are associated with low cancer risk, this does not constitute overwhelming evidence for any real protective effect, for the differences are not great. A moderate relative risk (for example, a twofold difference in cancer risk between the bottom third and the top third of some population with respect to carrier-bound retinol in the blood or β -carotene in the diet) might reflect a

truly protective effect, but it might arise indirectly merely because carrier-bound retinol levels are correlated with some other factor that is a real determinant of cancer risk. Even more plausibly, perhaps people who like to eat a lot of vegetables get both β -carotene and some truly protective factor from them, or tend to avoid some real cause of cancer in meat or other foodstuffs of which they consume less. These difficulties of interpretation can be reduced by seeking epidemiological evidence for a protective effect in many different studies in different types of population but, even if there is a medically significant but numerically moderate protective effect of dietary β -carotene, and even if these epidemiological studies nearly all estimate it reasonably accurately, we may still not be sure that they have done so unless the effect can be reproduced experimentally in randomly controlled prospective intervention studies among humans.

Blood retinol and cancer avoidance

The differences between total retinol and carrier-bound retinol concentrations in human blood are often smaller than the random error of measurement in blood retinol estimation, and so we shall not distinguish carefully between them in some of what follows.

In general, because of the wide range of blood levels of retinol in countries where vitamin A deficiency is common, any true protective effect may show up most strikingly in malnourished populations. However, the relevance of carrier-bound retinol and of β -carotene can be assessed independently only in well-nourished populations where variation in β -carotene intake has little effect on carrier-bound retinol levels and so the blood levels of the two factors are largely uncorrelated.

Prospective serum studies: Retinol can be reasonably stable in stored blood samples, and two studies^{31,32} have analysed the total retinol levels in stored samples of blood taken some years previously for unrelated research purposes from American and British people who were apparently free of cancer at that time. In both studies, people with total blood retinol levels near the top of the normal range were at significantly lower risk of developing cancer over the subsequent few years than those with levels near the bottom of the normal range (Table 1). But, although the apparent protective effects are statistically significant, so few cancers were observed that there is substantial statistical uncertainty in the magnitude of the effect. (Anyone who has been sitting for a few years on a few thousand blood samples might consider matching them up with the local cancer registry or with official death statistics, as long as it is still possible to measure the retinol and, ideally, the RBP and β -carotene in them.)

Against this, one preliminary observation which, if confirmed, suggests that retinol is not importantly protective relates to oral contraceptive (OC) use. OC users have substantially elevated levels of blood retinol^{14,15}, but the aggregated evidence from three studies of OC use in relation to the onset of various diseases, including all cancers, does not suggest that any material reduction in cancer risk is associated with 5-9 or even 10-14 years of continuous use of OCs. Tumours of the breast, endocrine, reproductive and genital organs are excluded (as these are likely to be directly dependent on hormonal factors or on sexual activity) and smoking habits are standardized for (since OC users tend to smoke more than non-users). A non-significant deficit of remaining tumour onsets among OC users in the Oxford (M. P. Vessey, personal communication and ref. 82) and Royal College of General Practitioners (C. R. Kay, personal communication and ref. 83) studies of a total of 60,000 women is counterbalanced by an apparent excess of skin cancer onsets among OC users in the American Nurses Health Study of 120,000 women (F. E. Speizer, personal communication; unfortunately, these American data could not be corrected for any hypothetical biases due to a correlation of sunbathing with OC use). Lack of any protective effect among long-term OC users would be quite strong evidence that blood retinol is not importantly protective, unless the rise in blood retinol among

Table 2 Case/control studies of blood retinol or β -carotene

Country	Mean values (mg l ⁻¹)		Substance	No. and site(s) of cancers
	cases	controls		
India ³³	0.31	0.48	Retinol	555 oral
India ³³	0.68	0.85	β -carotene	555 oral
Pakistan ³⁴	0.22	0.40	Retinol	203 oral
Pakistan ³⁴	0.40	0.62	β -carotene	203 oral
East Africa ³⁵	—*	—*	β -carotene	17 naso-pharyngeal
United Kingdom ³⁶	0.55	0.59	Retinol	35, many sites
United Kingdom ³⁶	0.83	0.92	β -carotene	35, many sites
United Kingdom ³⁷	0.46	0.66	Retinol	28 lung
United Kingdom ³⁸	0.47	0.60	Retinol	26 lung
United Kingdom ³⁸	1.05	1.26	β -carotene	26 lung
United Kingdom ³⁹	0.84	1.34	'Vitamin A'†	245, many sites
United States ⁴⁰	0.25	0.48	Retinol	51 gastro-intestinal
United States ⁴⁰	1.4	2.0	β -carotene	51 gastro-intestinal

* 'Carotene significantly lower' (in cases than in controls), but quantitative information not provided.

† Analytical methods not specified.

OC users is somehow due to a decrease in the ease with which peripheral cells can utilize retinol (or is in some other way compensatory).

Further large prospective studies of blood retinol are still needed. If they confirm the inverse association with cancer risks then identification of various intrinsic or extrinsic determinants of blood retinol will be necessary before it can be decided whether retinol *per se* is protective. Ideally, one might manipulate the determinants of blood retinol in a large randomized trial of cancer prevention, but as this may be impracticable (depending on the nature of these determinants) even uncontrolled observation, as in the above three studies of OC users, may provide useful clues.

Retrospective serum studies: Comparison of the total blood retinol levels of recently diagnosed cancer patients with the total blood retinol levels of control patients is easier and quicker than a prospective cohort study. Of the nine studies in India³³, Pakistan³⁴, East Africa³⁵, Britain³⁶⁻³⁹, or the United States^{40,41}, the eight which had at least a small control group³³⁻⁴⁰ all reported lower total retinol levels in cases than in controls (Table 2: all but one³⁶ of these eight are significant).

It may be relevant to the succeeding section to note that blood levels of β -carotene were also measured in some of these case/control studies. Among Indian, Pakistani and East African cases blood β -carotene levels were significantly lower than in controls. Unfortunately, β -carotene was measured in only three of the Western studies³⁸⁻⁴⁰, in each of which it was also somewhat lower in cases than in controls, even though no particular correlation between blood retinol and blood β -carotene levels need exist in developed countries.

The evidence provided by these eight case/control studies for either inverse association existing is not reliable, however, for in some studies there is no guarantee that the controls were appropriate and in all the studies the results might have been biased by a (hypothetical) 'nuisance effect' of depression of β -carotene or carrier-bound retinol by large, established cancers. There is still a need for large case/control studies in both developed and malnourished countries of carrier-bound retinol (and β -carotene) levels in people with small tumours that are unlikely to have had any material effect on diet or general metabolism (or in people who have been cured of malignant tumours some time previously with no obvious permanent damage to their digestive system or metabolism).

Dietary β -carotene and cancer avoidance

Turning from the blood to the diet, and shifting our emphasis from retinoids to carotenoids, a variety of independent epidemiological investigations have, in recent years, reported that cancer risks are inversely related to consumption of 'vitamin A'. This term embraces both retinol and the provitamin A carotenoids; but the total amount of 'vitamin A' in human diets from β -carotene (plus other carotenoids) often exceeds that from retinol.

In many cultures one particular plant or group of plants accounts for the majority of the dietary β -carotene: for example, carrots in North America⁴², yellow-green vegetables in Japan⁴³, red palm oil in West Africa⁴⁴ and dark green leafy vegetables among Singapore Chinese⁴⁵. Many epidemiological studies have investigated the relationship between β -carotene and cancer almost by accident, because dietary questionnaires designed for more general purposes happened to ask about the main local β -carotene source and it was realized only later that the lower cancer risks found among users of this particular foodstuff might be due to its β -carotene content. The correlation between the average β -carotene intake rates of different individuals over the past few years and their intake as assessed by such questionnaires cannot be high. So, if the observed 1.5- or two-fold relative risks (RR) of cancer typically observed when comparing low and high β -carotene intake groups (as assessed by dietary questionnaires) are indeed due to the β -carotene intake, the relative risk comparing true low with true high intake must presumably be somewhat more than

twofold, and might in principle be substantially greater. The results of these studies (5 prospective and 15 retrospective) are summarized in Table 3, and discussed below.

Prospective dietary questionnaires: All the Japanese in certain administrative areas were asked⁴³ to complete an extensive dietary questionnaire in 1965, and 265,118 did so. Over the next decade 807 died of lung cancer, yielding an age/sex/smoking-standardized relative risk of 1.4 ($P < 0.01$) for those who ate vegetables containing more than 1,000 IU of β -carotene per 100 g 'less than daily' with those who ate them 'daily'. Significant negative associations with daily use of these vegetables were also reported for stomach and prostate cancer, but the findings for other cancers were not reported. These yellow-green vegetables account for more than half of the β -carotene in the Japanese diet.

8,278 Norwegian men who had replied to postal questionnaires about both smoking (in 1964) and diet (in 1967) were followed for 5 years⁴⁶. A 'vitamin A index' (chiefly, in fact, a β -carotene index, since it included carrots and green vegetables but not liver or vitamin pills) was devised, low values of which were predictive of lung cancer incidence ($RR = 3$ for people in the bottom third of the population with respect to the 'vitamin A index' compared with people of similar age, sex and smoking habits in the top two-thirds; $P = 0.01$, but based on only 36 cases). A total of some 40,000 people completed the same questionnaire in 1966-67 in Norway or Minnesota⁴⁷, and the 10-year follow-up of these is being analysed, with apparently promising results (E. Bjelke, personal communication).

No other large prospective studies of either dietary β -carotene or preformed vitamin A have been published. It would be helpful if anyone who can link dietary questionnaires administered for any reason some years ago to subsequent cancer mortality (or, preferably, incidence) would do so, if they have reasonable estimates of consumption of at least the one or two main local β -carotene-containing plants plus at least 100,000 person-years of follow-up. For example, in Finland the detailed dietary questionnaires administered by the Research Institute for Social Security during 1966-72 to a geographically defined sample of 20,000 adults could be linked to the few hundred subsequent cancer onsets. Likewise, the Japan-Hawaii cancer project has by now sufficient data to provide some useful prospective information, although in Hawaii, as in Japan⁴³, one must interpret with particular caution correlations between any dietary habit and any cancer if both are subject to substantial trends, for artefacts will arise if the more modern members of society are more affected by both trends than are the more conservative members. Finally, dietary questionnaires have been administered to various large populations for research into cardiovascular disease, or for nutritional research, and these too might be exploitable. (Preliminary analyses of the 19-year follow-up of 2,000 people in Chicago who completed a 28-day dietary diary in 1959 indicate an inverse association of total cancer risk with β -carotene intake but not with retinal intake: R. Shekelle, personal communication.)

Retrospective dietary questionnaires: These involve asking people with cancer ('cases') and people without cancer ('controls') what vegetables, vitamin pills, etc. they normally ate in previous years, from which can be estimated⁴⁸ the relative risk of cancer in people with low, as opposed to high, β -carotene intake (or retinol intake—a nonsignificantly lower risk has been reported among users of vitamin A pills^{49,50}). One general difficulty is that feeling ill or knowing they have cancer may cause people to describe their past diet somewhat differently from how they would have described it if they had been healthy controls. If this bias is not substantial, the enormous advantage of the retrospective method is that it is quick (you do not have to wait for 10 years to see who develops the disease) and much less work (one questionnaire per case plus one per control), which allows more patients with cancer to be studied with consequently less random error in the final relative risk estimation. Although simple statistical methods⁴⁸ exist for correcting for age, sex, smoking and any dietary confounding variables

before calculating a test for a trend in the relative risk of cancer with respect to β -carotene or retinol intake, unfortunately investigators have often failed to use them or have used them but have failed to test for a *trend*, on only 1 degree of freedom.

At Roswell Park Memorial Institute in the United States, a few thousand cancer patients and a few thousand controls completed dietary questionnaires during 1957–62. Milk and some but not all other dietary sources of retinol were assessed, and carrots and some other major sources of β -carotene were assessed. A 'vitamin A' estimate was constructed for each person, which chiefly reflected provitamin A (β -carotene) intake, since on average five-eighths came from carrots (as β -carotene) and some more came from other vegetable sources of β -carotene⁴². Subjects were then classified as high, middle or low with respect to this estimate. When the 292 lung cancer patients were compared with selected controls⁵¹, the age/sex/smoking-standardized cancer risks relative to the high intake group were 1.0, 1.5 (middle) or 1.7 (low), and a test of the published data for an age/sex/smoking-standardized trend yields $P < 0.001$. An IARC case/control study of 233 Singapore Chinese lung cancers⁴⁵ indicated a $P < 0.01$ twofold protective effect of eating twice or more per week 'dark green leafy vegetables' (for example, kale), which are a major source of β -carotene for these people, but a similar case/control study⁵⁰ of 104 British lung cancers indicated no protective effect ($RR \approx 1.0$) of dietary 'vitamin A'. Returning to Roswell Park, the 569 bladder⁴², 122 oesophagus⁵² and 421 larynx⁵³ cancer patients were compared with selected controls. In seven sub-groups of decreasing (largely pro-) vitamin A intake, the risks of bladder cancer relative to those of the highest intake group were 1.0, 1.2, 1.4, 1.2, 1.7, 1.9 and 2.1, and a test of these published data for a trend again yields $P < 0.001$. Among the 122 oesophageal cancer patients divided into three sub-groups by dietary β -carotene intake relative risks of 1.0, 1.6 and 1.9 were found ($P = 0.03$ for trend), while among the 421 male larynx cancer patients divided into four dietary sub-groups, relative risks of 1.0, 2.1, 1.9 and 3.0 were found ($P < 0.005$ for trend) after adjustment for age, tobacco and alcohol. (Similar adjustment of the data for 581 oral cancers is necessary before the twofold protective effect seen in an unadjusted analysis⁵⁴ can be interpreted.)

The Roswell Park questionnaires comprise the largest set of case/control data available anywhere, and it is to be hoped that a single analysis of all cancers separately and together in relation to all controls will eventually become available in which trends in relative risks with respect to fine subdivisions of β -carotene and perhaps of retinol intake will be tabulated and standardized for finer subdivisions of age and smoking than heretofore.

There are also several studies relating gastrointestinal (GI) cancer risks to low vegetable intake rates but in all of these it is possible that vegetable components other than β -carotene (or dietary components other than the vegetables) account for the apparent protective effects in the GI tract.

Case/control studies of GI cancer have been reported from Hawaii^{55,56} and Israel^{57,58}, but in neither case were analyses reported with respect to specific estimates of β -carotene or retinol intake. Recently, the Israeli data on 406 cases and 812 controls have been re-analysed⁵⁹ with respect to an estimate of total dietary β -carotene for each case and for each control based on separate replies about more than 200 foodstuffs. There was no significant correlation between the risk of GI cancer and the estimated daily dietary carotene intake here (or in preliminary analyses of the Roswell Park data on GI cancers: S. Graham, personal communication). The only other study of GI cancer which has sought but not found any apparent protective effect whatever is the IARC case/control study of oesophageal cancer in French males⁶⁰, where the mean β -carotene (and retinol) intake was the same for the 200 cases as for the controls. By contrast, the IARC studies on oesophageal cancer in Iran⁶¹ indicate that the high-incidence areas were characterized by a diet extremely poor in green vegetables, and the main dietary finding in the subsequent IARC case/control studies within

high-incidence villages of several dozen dietary variables was a systematic lack of fresh fruit and vegetables among the cases⁶², although specific estimates of β -carotene (and retinol) intake in cases and controls were not published.

At other GI sites four separate case/control studies have been reported⁴⁷ each of which shows an age/sex-standardized relative risk of cancer that was statistically significantly greater than unity in people whose 'vegetable index' was lower, as opposed to higher, than normal. (Note, however, that only two control series, one in Norway and one in the United States, were interviewed.) The findings in these studies were⁴⁷:

(a) Norway,	228 stomach cancers:	$RR = 1.7$
(b) United States,	83 stomach cancers:	$RR = 1.5$
(c) Norway,	278 colorectal cancers:	$RR = 1.4$
(d) United States,	373 colorectal cancers:	$RR = 1.4$

In Liverpool, UK⁶³, a 1.6-fold relative risk of GI cancer ($P < 0.01$) was found if green vegetables were eaten less than weekly. Among vegetarians⁶⁴, a nonsignificant twofold relative risk was likewise found for colorectal cancer in Seventh Day Adventists who eat green, leafy vegetables 'less than' compared with 'at least' weekly. In India, it is reported⁶⁵ that studies of vegetarian sects consistently reveal smaller GI cancer risks, but no supporting data or references are provided. (Studies relating β -carotene intake to cancer risks within vegetarian sects would be free of confounding by associations with meat consumption.) Finally, in case/control studies⁶⁶ in a total of 521 gastric cancer patients in four countries, a lower intake of fresh fruits and vegetables in cases than in controls is indicated in the discussion, but the control data are not given (nor is the percentage of the controls who themselves had some other form of cancer).

Future research directions

Laboratory research: Most epidemiologists are well aware that the human evidence thus far available (Tables 2, 3) falls short of proof that increasing the average β -carotene intake would be of any material benefit in rich countries (or even in poor ones). *A fortiori*, the foregoing epidemiological studies may fail to interest many laboratory research workers whose habit is to trust only striking differences in well-controlled experiments. However, the perfect may well be the enemy of the relevant, for the detailed understanding of mechanisms which is slowly emerging from laboratory investigations of a few particular chemical, viral or physical agents (for example, SV40, phorbol esters or 2-acetylaminofluorene) may be of little immediate value if few human cancers are in reality due to these particular agents. Although in the long term fundamental research into any area of cell biology may prove relevant to the prevention or treatment of cancer, progress towards important preventive measures may be more rapid if, in addition to fundamental research, laboratory workers take the inevitably inconclusive leads provided by epidemiologists reasonably seriously (and vice versa, of course), one ultimate aim of this collaboration being to specify a few promising hypotheses in sufficient detail to make practicable their randomized evaluation among several thousand apparently healthy people for 5 or 10 years. Several laboratory studies are now possible that would test the plausibility of the more obvious mechanisms whereby β -carotene might be beneficial.

First, and most obvious, is to repeat the animal experiments in which important anticancer effects of β -carotene have been reported^{26–29}. Mechanistically, it would be interesting to discover whether dietary (not topical) β -carotene, started some weeks after weak (not massive) initiation of animals with DNA-binding carcinogens, has any material inhibitory effects on the later stage(s) of carcinogenesis, especially in tissues where dietary β -carotene is known to accumulate⁶⁷ or to cause accumulation of retinol⁶⁷.

Second, does β -carotene itself have any of the many direct effects on cultured cells or tissues that retinoids have? Such effects are relatively quick to obtain and easy to seek, although a variety of cell culture conditions and β -carotene vehicles may

Table 3 Questionnaire studies of cancer in relation to intake of some β -carotene-rich vegetable(s), or of 'vitamin A'. Note that 'Vitamin A' was estimated from a limited range of dietary components, chief among which in most of these studies has been the β -carotene in certain vegetables

Country	Relative risks (lower: higher intake)	No. of cancers	Subdivision by
Norway ⁴⁶	2.6: 1	36 lung	'Vitamin A'
Norway ⁴⁷	1.7: 1	228 stomach	Vegetables
Norway ⁴⁷	1.4: 1	278 colorectal	Vegetables
Minnesota, USA ⁴⁷	1.5: 1	83 stomach	Vegetables
Minnesota, USA ⁴⁷	1.4: 1	373 colorectal	Vegetables
Buffalo, USA ⁵¹	1.7: 1.5: 1	292 lung	'Vitamin A'
Buffalo, USA ⁴²	2.1: 1.9: 1.7: 1.2: 1.4: 1.2: 1	569 bladder	'Vitamin A'
Buffalo, USA ⁵²	1.9: 1.6: 1	122 oesophagus	'Vitamin A'
Buffalo, USA ⁵³	3.0: 1.9: 2.1: 1	421 larynx	'Vitamin A'
Singapore ⁴⁵	2.2: 1	233 lung	Green vegetables
United Kingdom ⁵⁰	0.7*: 1.3: 1.4: 1.1: 1	104 lung	'Vitamin A'
Israel ⁵⁹	1.1: 1.1: 1	406 gastrointestinal	β -carotene†
France ⁶⁰	1.0: 1	200 oesophagus	β -carotene
United Kingdom ⁶³	1.7: 1.1: 1	514 gastrointestinal	Green vegetables
US vegetarians ⁶⁴	2.2: 0.9: 1	41 colon	Green vegetables
Iran ⁶²	1.7: 1	344 oesophagus	Raw vegetables
Japan ⁴³	1.3: 1	611 male lung	"green/yellow‡" vegetables
Japan ⁴³	1.5: 1	196 female lung	
Japan ⁴³	2.0: 1	63 prostate	
Japan ⁴³	1.1: 1	1823 stomach	

Note: In addition, a US prospective study in 1948–54 reported an inverse association (risk ratios 2.2:1.6:1) with all-causes mortality⁸⁴.

* Bottom quintile based on only seven cancers, and so unreliable.

† Relative risks estimated from simultaneous regression on both retinol and β -carotene intake.

‡ Defined as any which contained over 1,000 IU of carotene per 100 g.

have to be studied. The weak effects already reported⁶⁸ could be pursued.

Third, although much of the enzymic oxidation of β -carotene to retinal takes place in the intestine, is there sufficient oxidation at other sites in humans to provide an appreciable source of localized retinol, or some other substance with retinoid-like activity? Dioxygenase, the enzyme which can oxidize β -carotene to retinal, has been demonstrated chiefly in the intestine, but it is also present in the liver and, to a lesser extent, the kidneys⁴. In which human tissues can β -carotene be converted into retinol? (When mice are treated with β -carotene deposits of it accumulate in many tissues, and deposits of retinol also accumulate in the lung and skin⁶⁷.)

Fourth, β -carotene is among the most efficient substances known for quenching the excitation energy of singlet oxygen^{24,25,69} and also for trapping certain organic free radicals⁷⁰. It has been suggested that anticancer effects (albeit perhaps on only the 'early' stage(s) of carcinogenesis) may be produced by deactivation of these chemical species⁷¹ or by otherwise preventing the lipid peroxidation or other oxidative damage which they might cause^{72–75}. This suggestion could, of course, be tested if people with a deficient ability to 'quench' singlet oxygen or to 'trap' certain types of free radical could be identified and their cancer onset rates monitored. If either chemical species is relevant, can realistic physiological concentrations of β -carotene, or any other carotenoids, effect any material reduction in the amounts of lipid peroxidation or other cellular damage caused by free radicals or by unquenched singlet oxygen, and is such oxidative damage importantly involved in carcinogenesis?

Fifth, both in poor and in rich countries, what are the chief dietary and other determinants of the differences that now exist between the blood levels of carrier-bound retinol and of β -carotene in different people? To what extent, for example, do they depend on dietary β -carotene, intestine-derived retinol, other micronutrients, or hormonal status? Comparing poor countries with each other⁷⁶, there is a correlation between dietary (pro-) vitamin A and blood retinol; but, as one would expect, the relationship seems to reach a maximum and then flatten off. Both oral contraceptives and pregnancy strongly affect blood retinol levels^{14,15}; what of menarche, menopause and dietary or medicinal hormones? Finally, diseases of the kidney¹⁶, thyroid^{5,16} and pancreas⁵ have been reported to cause increases or decreases in blood retinol or β -carotene; how do

they do so, and which other diseases (especially of hormone-secreting tissues) do so?

This list of laboratory questions could doubtless be extended. The general point is that the evidence linking β -carotene intake to cancer risk in man is already sufficiently consistent for laboratory workers to start to seek plausible explanations for it, bearing in mind the superficially paradoxical fact that dietary β -carotene may not affect serum retinol very much in developed countries.

Observational epidemiology: There are some populations (notably in parts of West Africa and Brazil) within which β -carotene intake varies enormously, due to the use of red palm oil. If β -carotene is really protective this wide variation should produce quite a striking variation in the relative risk of cancer, and so the Imperial Cancer Research Fund are supporting a case/control study (with J.D.B. and Dr. A. Kalache) in a part of Brazil where red palm oil is used by some, but not by all. Apart from this, and a general refining of questionnaires and statistical analyses to pick out the separate effects of β -carotene and retinol more specifically, we need more rigorous serum epidemiology to investigate the predictive value of serum retinol and β -carotene retrospectively or, preferably, prospectively. (Repeat measurements of retinol and β -carotene a year or more apart on a number of individuals to estimate within-person and between-person variability is a prerequisite for proper interpretation of such studies.) Future serum-based case/control studies are needed both in developed countries and in populations where hypovitaminosis A is endemic, but these should be large and should exclude artefactual effects of established tumours on blood levels (via effects on recent appetite or on general metabolism), perhaps by restricting interest to pre-cancerous lesions, to small non-digestive organ tumours, or to small tumours which have been cured without much trauma a year or more previously. Better, however, would be truly prospective studies, either starting *de novo* or utilizing blood samples taken years ago for other purposes but stored in conditions that still allow estimation of retinol and, ideally, β -carotene. Also, the question of whether prolonged oral contraceptive use reduces cancer risks should be pursued, as this may help determine whether any inverse association between blood retinol and cancer is due to a true protective effect of blood retinol, or merely to a secondary association (in which case, with what?). Likewise, any other extrinsic or intrinsic determinants of blood

retinol deserve epidemiological study. Finally, to help elucidate mechanisms, cancer onset rates might be monitored among small populations (if these can be found) with chronic metabolic defects which markedly elevate or depress blood levels of unbound retinol, carrier-bound retinol or β -carotene, or among people deficient in the β -carotene-oxidizing enzyme dioxygenase, in some of the normal mechanisms for quenching singlet oxygen, or in the enzymes for recognizing or utilizing retinol in peripheral tissues.

Intervention studies: In view of the general association previously observed between low vegetable use and cancer risks, it must be assumed that future, perhaps more specific, observational studies of serum β -carotene and retinol levels are in aggregate likely to continue to yield some inverse association with cancer risks, although the magnitude of the relative risk that will emerge cannot be predicted. If it is large then a clear need for randomized intervention studies will arise, but even if it is only moderate there will be no reliable way other than by randomized intervention to decide whether it reflects a cause-and-effect relationship. However, although analogy with the effects of giving up smoking suggests that inhibition of the late stages of human neoplastic progression should yield measurable effects within about 5 years, there may be little benefit apparent during the first 2 or 3 years of such a trial. Consequently, if randomized intervention trials are eventually to be undertaken then the sooner they are started the better, as long as they test β -carotene (preferably at as high a daily dose as possible without significant skin yellowing) rather than retinol, in the hope of detecting either any benefits of intestine-derived retinol or any direct benefits from circulating or adipose tissue β -carotene.

Testing β -carotene has the advantage of avoiding the risk of hypervitaminosis A arising among a few subjects who choose to overdose themselves⁵, for 'as far as is known hypercarotenosis in all its forms is entirely harmless'⁹ (and there is no evidence that the one man reputed to have died of carrot juice actually did so⁷⁷). Indeed, the joint FAO/WHO Expert Committee on Food Additives⁷⁸ formally estimated the acceptable daily intake of β -carotene for a 70 kg adult to be 350 mg per day! Since the normal range of human intake of this biologically active substance is only a few mg per day, this committee showed a remarkable faith in their ability to extrapolate from apparently complete safety in a few rats to safe levels in large human populations, but amounts of one or two hundred mg per day are regularly prescribed for the treatment of erythropoietic protoporphyria without causing hypervitaminosis A, liver dysfunction or any other apparent ill effect^{8,79-81} and large amounts are also consumed by many West Africans and Brazilians who prepare food with red palm oil. (Ghanaian liver retinol and blood carotene levels are grossly elevated, but their median blood retinol is normal⁴⁴.)

Long-term randomized intervention studies with β -carotene might be worthwhile on almost any category of persons at high enough risk of cancer for several dozen occurrences, recurrences or progressions to be anticipated in a practicable sized study group within a few years. However, among really high-risk people trials of synthetic retinoids may be preferred, despite the possibility of toxic effects. With β -carotene, perhaps the most interesting possibility is randomized intervention in the general population, or in those of the general population who seem to have less retinol or β -carotene than average by dietary questionnaires or by serum analysis. It may be most practical in such studies to restrict interest to people over 50 or over 60 years old, because in younger people (even if they smoke, which merely doubles their age-specific cancer risk) the cancer onset rate is so low that impractically large numbers would be needed to characterize the moderate benefits that might realistically be hoped for. Even if there is a really worthwhile benefit from β -carotene, it may require 5 years of study of many thousands of people aged over 50 years before clear evidence of benefit emerges, so that complicated follow-up and any other source of organizational complexity will have to be avoided if such trials are to be maintained for several years.

Conclusion

Blood retinol: In malnourished populations, increasing the dietary intake of β -carotene or retinol among people deficient in vitamin A can rectify the deficiency and cause blood retinol levels to rise, a procedure of such obvious medical value that it matters little from a public health point of view whether it also diminishes cancer incidence. Although this may likewise hold for that small minority of people in developed countries whose diet is seriously deficient in vitamin A, for most people in developed countries neither dietary β -carotene ('provitamin A') nor dietary retinol ('preformed vitamin A') have much effect on blood retinol. If the inverse association between blood retinol and cancer incidence that has recently been reported from small prospective studies in developed countries is confirmed in larger prospective studies it may well reflect a true cause-and-effect relationship. Even if it does, however, this will not suggest that increasing dietary vitamin A or provitamin A will reduce cancer incidence in a developed country for they have little effect on blood retinol. Rather it would suggest that the hormonal and other extrinsic factors which do influence blood retinol levels need to be discovered, and that when they have been the effects on total health of manipulating them will need to be evaluated rigorously in long-term trials.

β -carotene: When dietary β -carotene rather than blood retinol is considered, it appears that so many observational studies of dietary factors and cancer incidence have already indicated a slightly lower than average incidence of cancer among people with an above average intake of β -carotene that it is most unlikely that this association will disappear entirely with future observational studies, whether of blood or of dietary β -carotene. This inverse association may, of course, be an artefact, due merely to association of β -carotene ingestion with some truly protective dietary habit(s) or component(s), or to association of β -carotene ingestion with avoidance of some truly harmful dietary habit(s) or component(s). Equally, however, the inverse association between dietary β -carotene and cancer incidence may be due to a genuine protective effect of β -carotene against the onset of cancer. If so, perhaps it is the β -carotene that is absorbed unchanged from the intestine which is chiefly responsible for this protective effect rather than any retinol or retinal generated in the intestine from β -carotene, in which case this protective effect could not be produced by dietary retinol alone. Although various lines of laboratory and field research can be envisaged which might tip the balance of probability a little one way or the other as to whether β -carotene is or is not truly protective, it seems unlikely that this question will be decided unequivocally in the foreseeable future unless controlled trials are undertaken. Although the evidence thus far available is not compelling that β -carotene is truly protective against cancer (or, *a fortiori*, against total mortality) this is not a good reason to delay starting controlled trials because even if, as seems probable, a few truly protective agents do await discovery among the dozen or two dietary factors of current interest to the research community, compelling evidence may take decades to emerge without controlled trials.

Preventive measures, especially those which may be relevant over a long period to many millions of people, deserve peculiarly rigorous evaluation. Apart from reducing the amount consumed (or harmfulness) of tobacco, there are few lines of research or governmental action that offer any medium-term hope of achieving substantial reductions in total cancer incidence rates. Seeking extrinsic determinants of the blood levels of carrier-bound retinol is one, and investigating the anticancer effects of the carotenoids is another. In each case, whether or not a fruitful outcome is thought to be probable, the fact that it is at least plausible would justify pursuit of some of the suggested lines of research.

We thank Bruce Ames for suggesting the relevance of singlet oxygen and John Cairns for suggesting how to test this, Professor Saxon Graham and his co-workers Curtis Mettlin and Jim Marshall, Dr M. Mathews-Roth, Dr E. Bjelke, Dr R. Shekelle,

Dr R. Willson, Dr H. Cuckle, Professor B. Modan, Professor M. P. Vessey, Dr C. R. Kay and Dr F. E. Speizer, for permission to mention their unpublished results, and Mrs V. M. Godwin for repeatedly preparing the typescript.

1. Smith, J. E. & Goodman, D. S. *Fedn Proc.* **38**, 2504-2509 (1979).
2. Goodman, D. S. *Fedn Proc.* **38**, 2501-2503 (1979).
3. De Luca, L. M. *et al. Report of the International Vitamin A Consultative Group* (Nutrition Foundation, New York, 1978).
4. Olson, J. A. & Hayaishi, O. *Proc. natn. Acad. Sci. U.S.A.* **54**, 1364-1370 (1965).
5. Nieman, C. & Obbink, H. J. K. *Vita. Horm.* **12**, 69-99 (1954).
6. Crisp, A. H. & Stonehill, E. *Postgrad. med. J.* **43**, 721-725 (1967).
7. Robboy, M. S., Sato, A. S. & Schwabe, A. D. *Am. J. clin. Nutr.* **27**, 362-367 (American Society for Microbiology, Washington DC, 1980); *Oncology* (in the press).
8. Mathews-Roth, M. M. *et al. J. Am. med. Ass.* **228**, 1004-1008 (1974).
9. McLaren, D. S. in *Dermatology in General Medicine* 2nd edn. (eds T. B. Fitzpatrick, A. Z. Eisen, K. Wolff, I. M. Freedberg & K. F. Austen) (McGraw-Hill, New York, 1979).
10. Tlustohowicz, R. C. *et al. Research Report 56* (Storrs Agric. Experiment Station, University of Connecticut, 1979).
11. Varma, R. N. & Beaton, G. H. *Can. J. Physiol. Pharmac.* **50**, 1026-1037 (1972).
12. Vahlquist, A., Michaelsson, G. & Juhlin, L. *Acta derm-venereol.* **48**, 437-442 (1978).
13. Underwood, B. A., Siegel, H., Wiesell, R. C. & Dolinski, M. *Am. J. clin. Nutr.* **23**, 1037-1042 (1970).
14. Gal, I., Parkinson, C. & Craft, I. *Brit. med. J.* **2**, 436-438 (1971).
15. *Nutr. Rev.* **37**, 346-348 (1979).
16. Smith, F. R. & Goodman, D. S. *New Engl. J. med.* **294**, 805-808 (1971).
17. Boutwell, R. K. (consulting reviewer) *Selected Abstracts on Vitamin A in Cancer Biology* (International Cancer Research Data Bank, NCI, Bethesda, Maryland, 1979).
18. Schroder, E. W. & Black, P. H. *J. natn. Cancer Inst.* **65**, 671-674 (1979).
19. Sporn, M. B. & Newton, D. L. *Fedn Proc.* **38**, 2528-2534 (1979).
20. Peto, R. in *Origins of Human Cancer* (eds H. H. Hiatt, J. D. Watson & J. D. Winsten) (Cold Spring Harbor, New York, 1977).
21. Peto, R. *Proc. R. Soc. Lond. B* **205**, 111-120 (1979).
22. Rettura, G. *et al. 170th nat. Meet. Am. chem. Soc., Abstr.* **59**, (1975).
23. Felix, E. L., Cohen, M. H. & Loyd, B. C. *J. surg. Res.* **21**, 307-312 (1976).
24. Foote, C. S. in *Free Radicals in Biology* Vol. 2 (ed. Pryor, W.) 85-133 (Academic, New York, 1976).
25. Krinsky, N. I. in *The Carotenoids* (ed. Isler, O.) 669-716 (Birkhäuser, Basel, 1971).
26. Epstein, J. H. *Photochem. Photobiol.* **25**, 211-213 (1977).
27. Mathews-Roth, M. M. in *Proc. 12th int. Congr. Chemotherapists* (American Society for Microbiology, Washington DC, 1980); *Oncology* (in the press).
28. Seifter, E. & Rettura, G. *175th nat. Meet. Am. chem. Soc. Abstr.* **45**, (1978).
29. Dorogokupla, A. G. *et al. Zdravookhr. Kazakh.* **10**, 32-34 (1973).
30. Bagdon, R. E., Zbinden, G. & Studer, A. *Tox. appl. Pharmac.* **2**, 225-236 (1960).
31. Kark, J. D., Smith, A. H., Switzer, B. R. & Hames, C. G. *J. natn. Cancer Inst.* **66**, 7-16 (1981).
32. Wald, N. *et al. Lancet* **ii**, 813-815 (1980).
33. Wahi, P. N. *et al. Indian J. Path. Bact.* **5**, 10-16 (1962).
34. Ibrahim, J., Jafarey, N. A. & Zuberi, S. J. *Clin. Oncol.* **3**, 58-63 (1977).
35. Clifford, P. *Proc. R. Soc. Med.* **65**, 24-28 (1972).
36. Basu, T. K. *et al. Int. J. Vitam. Nutr. Res.* **44**, 14-18 (1974).
37. Basu, T. K. *et al. Br. J. Cancer* **33**, 119-121 (1976).
38. Atukorala, S. *et al. Br. J. Cancer* **40**, 927-931 (1979).
39. Capel, I. D. & Williams, D. C. *ICRS med. sci.* **7**, 361 (1979).
40. Abels, J. C., Gorbam, A. T., Pack, G. T. & Rhoads, C. P. *J. clin. Invest.* **20**, 749-764 (1941).
41. Cohen, M. H. *et al. Cancer Lett.* **4**, 51-54 (1977).
42. Mettlin, C. & Graham, S. *Am. J. Epidemiol.* **110**, 255-263 (1979).
43. Hirayama, T. *Nutr. Cancer* **1**(3), 67-81 (1979).
44. Dagadu, M. & Gillman, J. *Lancet* **i**, 531-532 (1963).
45. MacLennan, R. *et al. Int. J. Cancer* **20**, 854-860 (1977).
46. Bjelke, E. *Int. J. Cancer* **15**, 561-565 (1975).
47. Bjelke, E. in *Aktuelle Ernährungsmedizin* Suppl. 10-17 (Thieme, Stuttgart, 1978).
48. Mantel, N. & Haenszel, W. M. *J. natn. Cancer Inst.* **22**, 719-748 (1958).
49. Smith, P. G. & Jick, H. *Cancer* **42**, 808-811 (1978).
50. Gregor, A., Lee, P. N., Roe, F. J. C. *et al. Nutr. Cancer* **2**(2), 93-97 (1980).
51. Mettlin, C., Graham, S. & Swanson, M. *J. natn. Cancer Inst.* **62**, 1435-1438 (1979).
52. Mettlin, C., Graham, S., Priore, R., Marshall, J. & Swanson, M. *Nutr. Cancer* (in the press).
53. Graham, S. *et al. Am. J. Epidemiol.* (in the press).
54. Marshall, J., Graham, S., Mettlin, C. & Swanson, M. *Abstr. 13th Meet. Soc. Epidemiologic Research*, Minneapolis (1980).
55. Haenszel, W., Kurihara, M. & Segi, M. *J. Natn. Cancer Inst.* **49**, 969-988 (1972).
56. Haenszel, W. *et al. J. natn. Cancer Inst.* **51**, 1765-1779 (1973).
57. Modan, B. *et al. Cancer* **34**, 2087-2092 (1974).
58. Modan, B. *et al. J. natn. Cancer Inst.* **55**, 15-18 (1975).
59. Modan, B., Cuckle, H. & Lubin, F. (in preparation).
60. Tuyns, A. J., Pequinot, G. & Jensen, O. M. *Bull. Cancer* **65**, 59-64 (1978).
61. Cook-Mozaffari, P. *Nutr. Cancer* **1**(2), 51-60 (1979).
62. Cook-Mozaffari, P. *et al. Br. J. Cancer* **39**, 293-409 (1979).
63. Stocks, P. in *British Empire Cancer Campaign 35th a. Rep. 1957*, Part II, Suppl., 111-113 (1958).
64. Phillips, R. L. *Cancer Res.* **35**, 3513-3522 (1975).
65. Paymaster, J. C., Sanghri, L. D. & Gangadharan, P. *Cancer* **21**, 279-288 (1968).
66. Wynder, E. L. *et al. Cancer* **16**, 1461-1496 (1963).
67. Mathews-Roth, M. M., Crean, C. & Clancy, M. *Nutr. Rep. Int.* **16**, 419-423 (1977).
68. Mulson, R. A., DeFeo, D. & Weinstein, I. B. *Molec. Pharmac.* **16**, 569-578 (1979).
69. Foote, C. S., Chang, Y. C. & Denny, R. W. *J. Am. chem. Soc.* **92**, 5216-5218 (1970).
70. Packer, J. E. *et al. Biochem. biophys. Res. Commun.* **98**, 901-906 (1981).
71. Oberley, L. W. & Buettner, G. R. *Cancer Res.* **39**, 1141-1149 (1979).
72. Demopoulos, H. B., Pietronigro, D. D., Flamm, E. S. & Seligson, M. L. *J. envir. Path. Tox.* **3**, 273-303 (1980).
73. Ames, B. N., Holstein, M. C. & Cathcart, R. in *Lipid Peroxidation in Biology and Medicine* (ed. Yagi, K.) (Academic, New York, in the press).
74. Reddy, J. K., Azarnoff, D. L. & Hignite, C. E. *Nature* **283**, 397-399 (1980).
75. Shamberger, R. J. *J. natn. Cancer Inst.* **48**, 1491-1494 (1972).
76. Patwardhan, V. N. *Am. J. clin. Nutr.* **22**, 1106-1118 (1969).
77. Leitner, Z. A., Moore, T. & Sharman, I. M. *Proc. Nutr. Soc.* **34**, 44A (1975).
78. *W. d. H. th. Org. tech. rep. ser.* no. 557 (1974); *FAO Nutr. Meet. Rep. Ser.*, 1974, no. 54 (FAO, Rome, 1975).
79. Mathews-Roth, M. M. *et al. New Engl. J. Med.* **282**, 1231-1234 (1970).
80. Goerz, G. & Ippen, H. *Dt. med. Wsch.* **102**, 1051-1055 (1977).
81. Zaynoun, S. T. *et al. Brit. J. Derm.* **97**, 663-668 (1977).
82. Vessey, M. P. *et al. Lancet* **i**, 549-550 (1981).
83. Laydale, P. M. *et al. Lancet* **i**, 541-546 (1981).
84. Chope, H. D. & Breslow, L. *Am. J. publ. Hlth* **46**, 61-67 (1956).

ARTICLES

Earth's neodymium budget and structure and evolution of the mantle

Geoffrey F. Davies

Department of Earth and Planetary Sciences, McDonnell Center for the Space Sciences, Washington University, St Louis, Missouri 63130, USA

Neodymium and strontium isotopic data from rocks can be satisfied by a model in which the mantle contains chemical heterogeneities of many sizes which are less depleted in incompatible elements than the surrounding mantle, and in which the entire mantle is slowly mixed by convection.

NEODYMIUM isotopic abundance data have been used to constrain models of the structure and evolution of the Earth's mantle¹⁻⁵. According to some studies, only a fraction (one-third to one-half^{1,3-5}) of the mantle has been differentiated to form the continental crust, and the rest of the mantle is thought to retain its primitive Nd isotopic composition. This would favour models in which the mantle is compositionally layered, but not those in which the mantle is mixed on the large scale by whole-mantle convection. As our ideas on the evolution of the Earth will be strongly affected by which model is correct (geophysical constraints having proved inconclusive⁶⁻¹¹) so we need to establish the uncertainties of the isotopic constraints and the interpretations. Unfortunately, recent studies have only briefly considered the uncertainties of the constraints, they have assumed that only layered mantle models are viable, and the

complexities of calculations of the evolution of these models have obscured their sensitivity to uncertainties in the data.

This article examines the uncertainties of the Nd isotopic constraints; this is simplified by considering the present Nd budget of the Earth separately from the evolution of the Earth and of Nd. The mass of mantle which has been depleted to form the continental crust is found to be uncertain, and identifying this with the seismologically defined upper mantle is premature.

Some deficiencies of layered mantle models are also pointed out—the narrow range of Nd isotope data for ocean island basalts is not explained, and an important part of the rationale for layered models may be undermined by recent Nd data for kimberlite inclusions.

A model with a heterogeneous mantle and whole-mantle convection is proposed. It is pointed out that whole-mantle

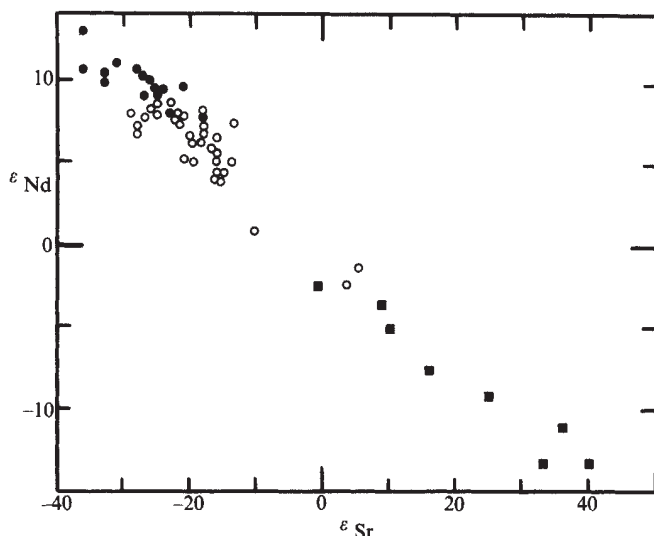


Fig. 1 Nd-Sr isotopic correlation for some oceanic rocks and kimberlite inclusions. Mid-ocean ridge basalt (MORB, ●) and oceanic island basalt (OIB, ○) data from ref. 2. Kimberlite inclusion (KI, ■) data from ref. 22. $\epsilon_{Nd} = 10^4(\alpha_{Nd}/0.511836 - 1)$ (refs 4, 12, 13). $\epsilon_{Sr} = 10^4(\alpha_{Sr}/0.7045 - 1)$ (refs 12, 14), where $\alpha_{Sr} = {}^{87}\text{Sr}/{}^{86}\text{Sr}$.

convection involves long overturn times and very long residence times, that plate tectonics must continuously generate heterogeneities, and that both mechanical and thermochemical factors may extend the survival times of heterogeneities to thousands of millions of years.

Nd isotope data

Variations in $\alpha_{Nd} = {}^{143}\text{Nd}/{}^{144}\text{Nd}$ produced by the decay of ${}^{147}\text{Sm}$ are conveniently measured in terms of ϵ_{Nd} (ref. 12), which measures parts in 10^4 deviations of α_{Nd} from a reference value that is based on measurements of the Juvinas meteorite¹³, and is assumed to represent an average chondritic value. Mid-ocean ridge basalts (MORB) have ϵ_{Nd} values near 10 (refs 12, 14–16), indicating that the MORB source (the oceanic mantle) is depleted in large-ion lithophile (LIL) elements relative to chondrites. Conversely, many young continental rocks have negative ϵ_{Nd} values^{17–20} indicating a complementary LIL-enrichment relative to chondrites. Island arc rocks tend to have $\epsilon_{Nd} \sim 8$, while oceanic island basalts (OIB) tend to have $\epsilon_{Nd} \sim 6$, but with some ranging to near zero^{2,12,14,16}. Continental flood basalts (CFB) tend to have ϵ_{Nd} values near zero²¹, which suggests they derive from a source which is primitive and chondritic¹⁴. A negative correlation between ϵ_{Nd} and ϵ_{Sr} has been found for oceanic rocks^{12,14–16}, where ϵ_{Sr} is the analogous measure of the Rb–Sr system¹² with $\epsilon_{Sr} = 0$ being defined by the point where the correlation crosses $\epsilon_{Nd} = 0$. This correlation is shown in Fig. 1 for a selection of MORB and OIB samples². Recently some kimberlite inclusion (KI) data have been obtained which extend this correlation through the origin²² (Fig. 1). Other continental rocks and island arc rocks often deviate from this correlation, and this can often be attributed to contamination by seawater or lower continental rocks of magma originally lying on the correlation line^{17,20,23–25}.

Nd in chondrites

Interpretations of the above data have been based on the assumptions that the average crust–mantle system (bulk Earth or BE) has chondritic Sm/Nd and Rb/Sr ratios, and that chondritic material will have $\epsilon_{Nd} = 0$. The first assumption is very difficult to justify with any rigour. For example, if chondrites derive from a relatively small number of parent bodies, then our sample may be significantly biased. Furthermore, the Earth may have sampled a slightly different population of material, because

many meteorites are believed to be derived from the asteroid belt, more than twice as far from the Sun as the Earth.

There are some limited data with which the second assumption has been tested. Jacobsen and Wasserburg²⁶ have reported Sm and Nd isotopic and abundance data for several chondrites and achondrites (including Juvinas). Their results imply $\epsilon_{Nd} = -1 \pm 1$ at present and $\epsilon_{Nd} = -2 \pm 0.5$ at 4,500 Myr ago, but it is not clear to what extent these discrepancies from $\epsilon_{Nd} = 0$ may be due to interlaboratory biases^{26,27}. (Note that these ϵ values are based on the old reference values^{12,13}, and not on the new ones defined by Jacobsen and Wasserburg²⁶.) About 50 isotope dilution measurements of Sm/Nd in chondrites^{28–32} give a mean ${}^{147}\text{Sm}/{}^{144}\text{Nd}$ of 0.196 ± 0.002 (see also refs 26, 27), compared with the Juvinas reference value of 0.1936 (ref. 13). If all chondritic material began with $\epsilon_{Nd} = 0$, this Sm/Nd range would generate a range of 2 ± 6 (1 s.d.) in ϵ_{Nd} , with the mean being $\epsilon_{Nd} = 2 \pm 2$ (2 s.d.).

Present Nd mass balance

Suppose that the mantle initially had mass M , Nd mass concentration C_0 and $\epsilon_{Nd} = \epsilon_0$. Suppose that later the mantle comprises several reservoirs with masses M_i , concentrations C_i and $\epsilon_{Nd} = \epsilon_i$, and the mass fraction $x_i = M_i/M$. Then the equations which express conservation of the masses of the mantle and Nd and the total amount of radiogenic Nd produced in the interval are

$$\begin{aligned} \sum x_i &= 1 \\ \sum x_i C_i &= C_0 \\ \sum x_i C_i \epsilon_i &= C_0 \epsilon_0 \end{aligned} \quad (1)$$

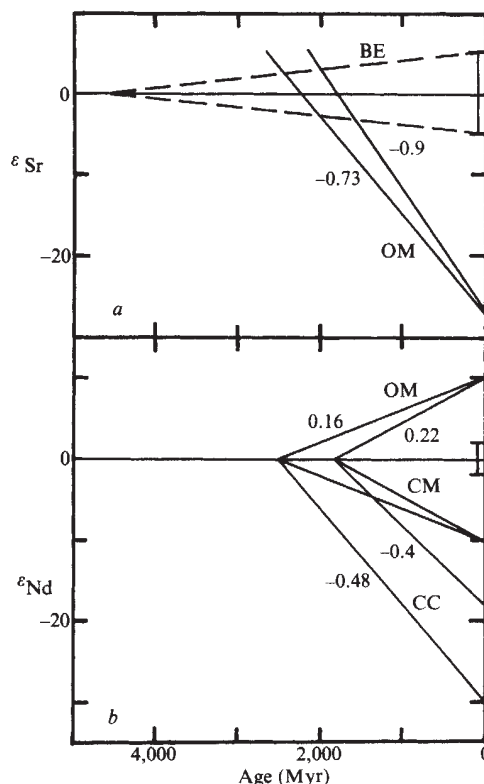


Fig. 2 Models of ϵ_{Sr} and ϵ_{Nd} evolution calculated from equation (1) and its Sr analogue ($Q_{Sr} = 0.0167 \text{ Myr}^{-1}$, ref. 4). *a*, Evolution of ϵ_{Sr} in the oceanic mantle (OM) relative to the bulk Earth (BE). OM lines are labelled with the Rb–Sr enrichment factor, f_{Sr} . Error bar on BE is the estimated uncertainty of the present BE value of ϵ_{Sr} (see text). $f_{Sr} = (\mu_{Sr}/0.0839 - 1)$, where $\mu_{Sr} = {}^{87}\text{Rb}/{}^{86}\text{Sr}$ (refs 4, 12, 14). *b*, Evolution of ϵ_{Nd} for oceanic mantle (OM), continental mantle (CM) and continental crust (CC) for mean differentiation ages of 1,800 and 2,500 Myr. Lines are labelled with values of f_{Nd} .

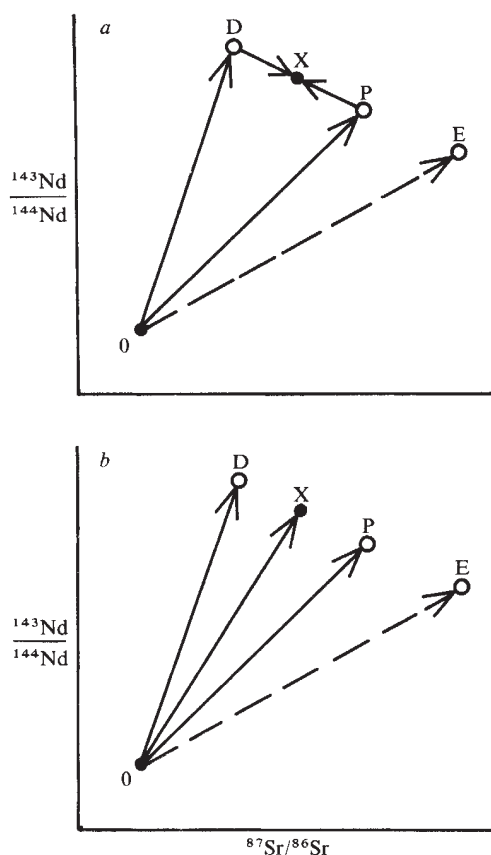


Fig. 3 Alternative explanations of the correlation in Fig. 2. *a*, Depleted (D) and primitive (P) sources evolve independently from 0. Mixing from D and P then yields the isotopic composition X. The presence of an enriched source (E), as suggested by kimberlite inclusion data²², implies that either E evolved independently of but colinearly with D and P, or that P results from mixing of D and E and is therefore not primitive. Straight mixing lines require the sources to have the same Sr/Nd ratio⁴⁶. *b*, All sources evolve independently but colinearly. This requires coherent fractionation of Sm/Nd and Rb/Sr: specifically, that sources have the same ratio of enrichment factors, f_{Nd}/f_{Sr} (refs 2, 53).

These relations are implicit in the treatments of Nd evolution and partitioning by O'Nions *et al.*³ and Jacobsen and Wasserburg⁴, but they did not separate the instantaneous mass balance from the evolution. In equations (1), the evolution is implicit in the present values of ϵ .

Some results of applying equations (1) are given in Table 1. Cases 1–4 are variations on Model 1 of Jacobsen and Wasserburg⁴ (which is similar to Case 3, Table 1): the present reservoirs considered are the continental crust (CC), depleted mantle (DM), and primitive mantle (the same as the initial mantle, but with a smaller mass). The parameters of the initial mantle (IM) and of CC were assumed, and those of DM were calculated, assuming $\epsilon_{DM} = 10$, consistent with Fig. 1 (ref. 4). The initial concentration of Nd, $C_0 = 0.9 \mu\text{g per g}$, is lower than that assumed by Jacobsen and Wasserburg⁴, and is based on thermal evolution models of the Earth^{33,34} which require uranium abundances in the mantle to be ~ 1.5 times chondritic: C_0 is also assumed to be 1.5 times chondritic, rather than twice chondritic⁴.

Case 1, Table 1, requires about half of the mantle to have been depleted ($x_{DM} = 0.48$). If ϵ_{CC} is assumed to be -30 rather than -20 , which implies an older crust, then about two-thirds of the mantle must have been depleted (Case 2). On the other hand, if $C_0 = 1.26 \mu\text{g per g}$ (ref. 4), then only one-third of the mantle need have been depleted (Case 3). The sensitivity of the mass

balance to ϵ_0 is illustrated by Case 4: almost the entire mantle must have been depleted if ϵ_0 is increased to $+4$.

There is increasing evidence that the sub-continental mantle is veined by LIL-element-enriched phases which were emplaced by metasomatic events into previously depleted mantle^{23,35–37}. The kimberlite-inclusion data in Fig. 1 suggest that $\epsilon_{Nd} = -10$ may be reasonably representative of this heterogeneous material. Its total abundance is difficult to estimate, but one constraint may come from continental heat flow data, from which it is inferred that $\sim 25 \text{ mW m}^{-2}$ of heat flux comes from below the upper 10–15 km of the crust³⁸: this is compatible with uranium and thorium abundances in the continental lithosphere up to ~ 10 times chondritic abundances (that is, U abundances up to $\sim 0.15 \mu\text{g per g}$). Jordan³⁹ points out that continental garnet lherzolite inclusions have K and other LIL abundances compatible with this degree of enrichment, and he summarizes the confluence of geophysical and geological data favouring the existence of an ancient, stable, LIL-enriched subcontinental mantle. Thus the subcontinental mantle may be an ancient reservoir with significant Sm and Nd enrichments.

Cases 5–8 of Table 1 explore the possible effect of an additional reservoir in the sub-continental lithospheric mantle (CM) by giving the Nd concentration in CM which would be required by depletion of the entire remaining mantle ($x_{CM} = 0.03$ corresponds to continental lithosphere $\sim 200 \text{ km}$ thick). Cases 5, 6 and 8 yield abundances from ~ 3 – 13 times chondritic ($0.6 \mu\text{g per g}$); only Case 7 clearly violates the constraint inferred from heat flow.

These results demonstrate that the fraction of mantle which has remained undepleted is subject to large uncertainties. Thus, the identification of the undepleted mantle with the lower mantle is speculative, and most of the mantle may possibly have been depleted.

Age of the continental crust

Because of the long half life of ^{147}Sm , ϵ_{Nd} varies approximately linearly with time in a closed Sm–Nd system, and can be written⁴:

$$\epsilon_{Nd}(t) = \epsilon_{Nd}(0) + Q_{Nd} f_{Nd} t \quad (2)$$

where $Q_{Nd} = 0.02474 \text{ Myr}^{-1}$ is a constant related to the decay constant, t is time and f_{Nd} is the Sm–Nd enrichment factor of the reservoir relative to Juvinas: $f_{Nd} = (\mu_{Nd}/0.1936 - 1)$ and $\mu_{Nd} = ^{147}\text{Sm}/^{144}\text{Nd}$. Jacobsen and Wasserburg⁴ have shown that equation (2) yields an upper bound, T , on the mean age of a reservoir if $\epsilon(-T)$, $\epsilon(0)$ and f_{Nd} are known.

The chronology of the continental crust is a major question in the mass-balance calculations (and in Earth history in general). Equation (2) shows explicitly how the evolution of reservoirs enters the mass balance calculations of the previous section. For example, Case 2 (Table 1), with $\epsilon_{CC} = -30$, would require the continental crust to be older, or to have a more negative f_{Nd} , or both, than Case 1, with $\epsilon_{CC} = -20$.

Table 1 Nd mass balance

Case	Initial mantle ($x = 1$) C ($\mu\text{g per g}$)	ϵ	x	Depleted mantle ($\epsilon = 10$) C ($\mu\text{g per g}$)	Continental crust ($x = 0.0055$, $C = 26$) ϵ	Continental mantle ($x = 0.03$, $\epsilon = -10$) C ($\mu\text{g per g}$)
1	0.9	0	0.48	0.60	-20	—
2	0.9	0	0.64	0.68	-30	—
3	1.26	0	0.34	0.84	-20	—
4	0.9	4	0.93	0.75	-25	—
5	0.9	0	0.97*	0.53	-20	8
6	0.9	0	0.97*	0.60	-30	5
7	1.26	0	0.97*	0.74	-20	14
8	0.9	4	0.97*	0.71	-20	1.6

* Assumed.

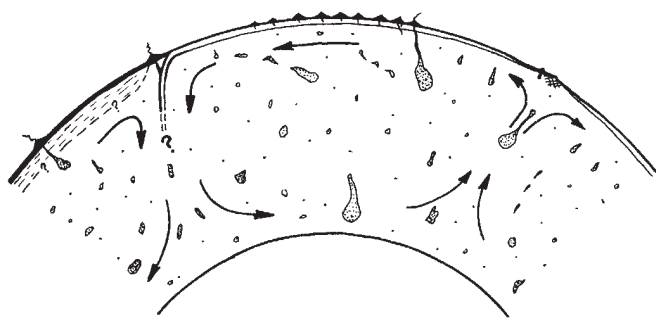


Fig. 4 Some aspects of the heterogeneous mantle models. Blobs represent chemical heterogeneities which are less LIL-depleted than the surrounding mantle. Heterogeneities are of many sizes, but their geometry is unknown; they may be generated at subduction zones and higher concentrations of radioactive heat sources would make them prone to diapirism; they would tend to be destroyed or dispersed by the shearing of convective flow.

Jacobsen and Wasserburg⁴ estimated the mean age of the crust-mantle differentiation using the Rb-Sr system in the oceanic mantle (OM). For OM, the present ϵ_{Sr} is about -27 (Fig. 1), and they estimate that f_{Sr} lies between -0.8 and -0.95, with -0.9 being their preferred value. The mean age, T_{Sr} , of OM can then be obtained from the Sr analogue of equation (2), and the graphical solution is shown in Fig. 2a: assuming that initially OM had $\epsilon_{Sr} = 0$, the result is $T_{Sr} \leq 1,800$ Myr. This is a rather small age for the depleted OM and its complementary continental crust (CC), and would favour Cases 1 and 3 ($\epsilon_{Nd} = -20$) rather than Case 2 ($\epsilon_{Nd} = -30$), as is shown in Fig. 2b assuming that $f_{Nd} = -0.4$ for CC⁴.

Jacobsen and Wasserburg's^{4,40} confidence in this estimate is based on the presumption that the OM is strongly depleted in Rb, so that f_{Sr} must be close to -1 and its absolute uncertainty cannot be large. However, Weaver and Tarney⁴¹, for example, estimate $f_{Sr} = -0.73$ for OM, which yields $T_{Sr} \leq 2,200$ Myr (Fig. 2a). Note that the bulk Earth value of ϵ_{Sr} is inferred from the correlation in Fig. 1 and is, therefore, uncertain by about ± 5 (and the correlation may not even preserve direct information about the initial ϵ_{Sr} of the Earth). When account is taken of such uncertainties in enrichment factors and ϵ values of various reservoirs^{3,4,41}, one must conclude that equation (2) and its Sr analogue do not yet give very strong constraints on the mean age of the crust-mantle differentiation (Fig. 2a).

On the other hand, the observations that the continents contain provinces which were stabilized by the late Archaean (~2,500 Myr ago), are of average crustal thickness (35-40 km) and which have little or no sediment cover lead to the conclusion that much of the continental crust had been differentiated by the late Archaean (see refs 42-44). Wise⁴² estimates that 90% of the continental crust had been differentiated by the end of the Archaean. This conclusion can be avoided if the Archaean crust had subsequently been thickened by underplating, but there is no recognized modern analogue of such a process, and it would have to be an extremely delicate process to affect broad areas uniformly without leaving recognizable effects at the surface. This conclusion would imply that the mean age of differentiation of the continental crust from the mantle is $\geq 2,500$ Myr.

If the mean age of the continental crust is as great as 2,500 Myr, and with estimates of f_{Nd} of CC ranging from -0.35 to -0.48 (refs 3, 4, 41, 45), then it is quite plausible for ϵ_{Nd} of CC to be -30, as assumed in Table 1. It follows that well over half of the mantle may be LIL-depleted. Note that a Sr mass balance may also be calculated^{3,4} as was done here for Nd, but the uncertainties for that calculation are at least as great: the mean Rb-Sr enrichment of the continental crust, in particular, is very uncertain.

The layered mantle model

Wasserburg and DePaolo¹ have advocated a two-reservoir mantle model in which the (seismologically defined) upper mantle is identified as depleted oceanic mantle and the lower mantle is identified as a primitive (chondritic) source of the continental flood basalts. This division is suggested by their mass balance estimates^{1,4} (see case 3, Table 1). In this model the correlation in the depleted quadrant of Fig. 1 comprising MORB, island arcs, OIB and CFB (not shown, but with $\epsilon_{Nd} \approx 0$), is interpreted as a mixing line between MORB and CFB. If that is true, then the clustering of the OIB data at $\epsilon_{Nd} = 6 \pm 2$ is puzzling: why do not all degrees of mixing occur? As no one seems to advocate a special intermediate source reservoir with $\epsilon_{Nd} = 6$, the clustering must be due either to the small data sample or to some unknown mechanism. If either of these is true for the OIB, then why not also for the CFB, clustering at $\epsilon_{Nd} = 0$, especially because the new KI data suggest that the correlation of Fig. 1 extends into the enriched quadrant?

There are two mechanisms which might yield a correlation like that in Fig. 1, and these are shown in Fig. 3: mixing between two reservoirs (Fig. 3a), and coherent fractionation with independent evolution of sources (Fig. 3b). As long as only two reservoirs were sufficient to account for the data, the mixing hypothesis was perhaps preferable. However, the advent of the KI data (Fig. 1) necessitates at least one more reservoir: an enriched KI endmember (E, Fig. 3). If these reservoirs have evolved independently and colinearly, then coherent fractionation is established and can be invoked for any other point on the correlation line. Alternatively, if all points between D and E result from mixing of D and E (requiring identical Sr/Nd (ref. 46)), then it would only be coincidence that the CFB have $\epsilon_{Nd} = 0$, the identification of the CFB source as primitive would be premature, and the primitive values of ϵ_{Sr} would be unknown. Furthermore, the geometry of the sources would have to be revised. In either case, the KI data make the case for the layered two-reservoir mantle model much less compelling.

Another important question about the layered mantle model is how the layers separate (note that the separation must not be complete, as lower mantle material is required to reach the surface). The mantle transition zone may effect this separation, either by the effects of phase changes or, more likely, by a change in intrinsic density⁶. If the upper mantle is separated in this way and has been steadily depleted, then, according to Jacobsen and Wasserburg⁴, the concentrations of some highly incompatible elements in crustal rocks would have decreased 10-fold in the past 3,500 Myr, and they claim there is no evidence for this. If, as they prefer, the depleted mantle has been steadily growing in volume, with a constant degree of depletion, then it would only be coincidence if its lower boundary coincided with the transition zone at present. If the boundary between depleted and undepleted layers does not coincide, or has not coincided, with the transition zone, then the existence of the boundary and its ability to prevent mixing become *ad hoc*: it is certainly possible, but there is no compelling geophysical evidence for another boundary in the mantle.

Heterogeneous mantle model

At this stage, the Nd data can be interpreted in terms of an alternative model of the mantle which is chemically heterogeneous up to the largest scales and in which the entire mantle is slowly mixed by convection. The geophysical evidence is equivocal on the question of the structure of mantle convection: both whole-mantle convection^{6,7,11} and convection in separate layers⁸⁻¹⁰ have been recently advocated on the basis of geophysical evidence, so geochemists must also consider these alternatives.

The type of model proposed is shown in Fig. 4. It is postulated that the average mantle is fairly strongly LIL-depleted ($\epsilon_{Nd} \approx +10$) and that a small volume fraction (perhaps 10-20%) comprises less depleted or enriched regions. These regions are depicted in Fig. 4 as blobs, but their geometry is unknown. They

could be remnants of primitive mantle ($\epsilon_{Nd} \approx 0$), but it is more likely that they are products of more complex differentiation histories involving both enrichments and depletions, resulting in a wider range of ϵ_{Nd} ($10 \geq \epsilon_{Nd} \geq -10$). If these differentiation events involved coherent fractionation, as at least some must have, according to the earlier discussion, then the linear array of Fig. 1 would result. The heterogeneities may range in size up to hundreds or even thousands of kilometres, as suggested by the largest oceanic volcanic hotspots (Iceland and Hawaii), and down to grain sizes³⁵⁻³⁷. The mantle heterogeneities must have existed through much of the Earth's history, as indicated by the mean age of the crust and mantle reservoirs (Fig. 2), and by inferred ancient metasomatic events²³.

Subduction zones must generate mantle heterogeneities, because subducting oceanic lithosphere is strongly stratified and it apparently remains mechanically coherent at least to depths of hundreds of kilometres. The upper mantle part of subducted lithosphere is presumably LIL-depleted and refractory: it may therefore be mechanically strong and persistent, while pieces of it may ultimately become incorporated into continental lithosphere^{39,47}. The crustal part is enriched in some LIL elements and hydrated⁴⁰: much of the enriched component may be incorporated into island arc magmas^{20,24,25}, but some probably escapes this process and returns to the mantle⁴³. Until the complexities of island arc geochemistry and of sialic crust generation are better understood, it is difficult to predict the character of such LIL mantle enrichments, but if they are significantly enriched in radioactive heat sources (U, Th, K), they would be prone to diapiric instability and uprising as mantle viscosity is a strong function of temperature. Such diapirs would be likely to mobilize and concentrate volatile and other incompatible elements, and thus might be the cause of mantle metasomatic events. They would be, in this model, the immediate sources of volcanic hotspots, although only a small fraction of all existing diapirs might reach the surface at a given time; many others might be disrupted by shearing before reaching the surface. Some of these features are shown in Fig. 4.

An important question is how the postulated heterogeneities could survive in a fully convective mantle, as the shearing in convective flow tends to smear out heterogeneities^{48,49}: the following arguments demonstrate the plausibility of heterogeneities surviving for thousands of millions of years. In a whole-mantle convection system, at present rates of seafloor spreading, it would take over 4,000 Myr for the mass of the mantle to pass within 100 km of the surface⁵⁰, and thus to undergo a significant melting and differentiation event; that is the mean residence time between major differentiation events is thousands of millions of years. For isolated more-viscous regions, shearing rates vary inversely as viscosity, so that refractory regions are likely to have low shearing rates. On the other hand, shearing rates of isolated less viscous regions are controlled by the viscosity of the surrounding mantle and thus will not be exceptionally large. In addition, the diapiric tendency of such regions, suggested above, will tend to enhance enriched heterogeneities.

A heterogeneous mantle is a natural outcome of the present tectonic regime, but the degree of chemical heterogeneity is difficult to estimate as it will be the result of a dynamic equilibrium

between the competing processes creating and destroying heterogeneities. This can be characterized in terms of the residence times of mantle material and survival times of heterogeneities. Residence times are clearly of the order of 10^9 years at present. Survival times depend on the rates of competing processes and these are difficult to estimate. The shearing rate can be characterized by the mantle overturn time which, at 30 mm yr^{-1} , is at least 300 Myr for the whole mantle. In a homogeneous fluid, heterogeneities can become very elongated by shearing in one overturn^{46,47}, but the mechanisms discussed above might easily extend the survival time well beyond one overturn time.

Because the Earth's heat flow was probably greater in the past, all processes would have been faster. Plate velocities are approximately proportional to the square of the heat flux⁵¹, but it is unlikely that the heat flux has been much more than twice its present value in the past 2,000 Myr (refs 33, 52). Thus residence times have probably been $> 1,000$ Myr for the past 2,000 Myr, and other rates would probably have changed in proportion.

O'Connell and Hager⁵⁰ have pointed out that residence times may be considerably less in the upper mantle than in the lower mantle because of the relatively rapid migration of ridges and trenches compared with convective velocities. Essentially, the more recently subducted material will have a greater chance of being returned to the surface by a passing ridge than deeper material. Thus this kinematic effect would tend to produce a stratification (but not necessarily layering) of the mantle in which the deeper material has been less differentiated than shallower material.

Conclusions

The Nd and Sr isotopic data clearly demand a mantle in which heterogeneities have existed for thousands of millions of years: the real question is the topology and distribution of the heterogeneities. In this sense the layered mantle model is a special case in which there are only vertical heterogeneities with sharp boundaries. Our heterogeneous mantle model is another type in which there is little or no mean vertical variation and in which the heterogeneities are isolated and have a range of chemical differences and a large range of sizes. The principal differences between these models lie in (1) the fraction of the mantle which has been strongly LIL-depleted, (2) the topology of the heterogeneities and (3) the existence of a reservoir of isotopically primitive material.

The principal question about the heterogeneous mantle model is probably the uncertainty of the survival time of the heterogeneities, while the principal question about the layered model may be the mechanism by which the layers are separated.

If some heterogeneities must exist within the layers of the layered mantle model, and if the heterogeneous model has some vertical stratification, then the distinctions between the models begin to blur. The real issue then is not a choice between two idealized models, but the accuracy with which the Earth's place in the spectrum of models can be determined. It is concluded here that no part of the spectrum of models discussed above can confidently be excluded at present.

I thank Don DePaolo, Gerry Wasserburg, Rodey Batiza, Larry Haskin and Keith O'Nions for constructive comments.

Received 31 July 1980; accepted 5 January 1981.

1. Wasserburg, G. J. & DePaolo, D. J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3594-3598 (1979).
2. Allègre, C. J., Ben Othman, D., Polve, M. & Richard, P. *Phys. Earth planet. Inter.* **19**, 293-306 (1979).
3. O'Nions, R. K., Evensen, N. M. & Hamilton, P. J. *J. geophys. Res.* **84**, 6091-6101 (1979).
4. Jacobsen, S. B. & Wasserburg, G. J. *J. geophys. Res.* **84**, 7411-7427 (1979).
5. DePaolo, D. J., *Geochim. cosmochim. Acta* **44**, 1185-1196 (1980).
6. Davies, G. F. *Geophysical. J. R. astr. Soc.* **49**, 459-486 (1977).
7. Hager, B. H. & O'Connell, R. J. *J. geophys. Res.* **84**, 1031-1048 (1979).
8. Richter, F. M. *J. geophys. Res.* **84**, 6783-6795 (1979).
9. Parmentier, E. M. & Turcotte, D. L. *Phys. Earth planet. Inter.* **17**, 281-289 (1978).
10. McKenzie, D. P. & Weiss, N. O. *Geophys. J. R. astr. Soc.* **42**, 131-174 (1975).
11. Jordan, T. H. *J. Geophys.* **43**, 473-496 (1977).
12. DePaolo, D. J. & Wasserburg, G. J. *Geophys. Res. Lett.* **3**, 249-252, 743-746 (1976); **4**, 465-468 (1977).
13. Lugmair, G. W., Scheinin, N. B. & Marti, K. *Earth planet. Sci. Lett.* **27**, 79-84 (1975).
14. DePaolo, D. J. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **43**, 999-1008 (1979).

15. O'Nions, R. K., Hamilton, P. J. & Evensen, N. M. *Earth planet. Sci. Lett.* **34**, 13-22 (1977).
16. Richard, P., Shimizu, N. & Allègre, C. J. *Earth planet. Sci. Lett.* **31**, 269-278 (1976).
17. Carter, S. R., Evensen, N. M., Hamilton, P. J. & O'Nions, R. K. *Science* **202**, 743-747 (1978).
18. Allègre, C. J. & Ben Othman, D. *Nature* **286**, 335-342 (1980).
19. McCulloch, M. T. & Wasserburg, G. J. *Science* **200**, 1003-1011 (1978).
20. Hawkesworth, C. J. & Vollmer, R. *Contr. Miner. Petrol.* **69**, 151-165 (1979).
21. DePaolo, D. J. & Wasserburg, G. J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3056-3060 (1979).
22. Menzies, M. & Murthy, V. R. *Nature* **283**, 634-636 (1980).
23. Menzies, M. & Murthy, V. R. *Earth planet. Sci. Lett.* **46**, 323-334 (1980).
24. DePaolo, D. J. *Science* **209**, 684-687 (1980).
25. Hawkesworth, C. J. *et al. Earth planet. Sci. Lett.* **42**, 45-57 (1979).
26. Jacobsen, S. B. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **50**, 139-155 (1980).
27. Lugmair, G. W. & Carlson, R. W. *Proc. 9th Lunar planet. Sci. conf.* 689-704 (1978).
28. Masuda, A., Nakamura, N. & Tanaka, T. *Geochim. cosmochim. Acta* **37**, 239-248 (1973).
29. Nakamura, N. *Geochim. cosmochim. Acta* **38**, 757-775 (1974).
30. Nakamura, N. & Masuda, A. *Earth planet. Sci. Lett.* **19**, 429-437 (1973).
31. Gast, P. W., Hubbard, N. J. & Wiesmann, H. *Proc. 2nd Lunar Sci. Conf.*, 1143-1163 (1970).

32. Evensen, N. M., Hamilton, P. J. & O'Nions, R. K. *Geochim. cosmochim. Acta* **42**, 1199–1212 (1978).
33. Davies, G. F. *J. geophys. Res.* **85**, 2517–2530 (1980).
34. Schubert, G., Stevenson, D. & Cassen, P. *J. geophys. Res.* **85**, 2531–2538 (1980).
35. Lloyd, F. E. & Bailey, D. K. *Phys. Chem. Earth* **9**, 385–416 (1975).
36. Frey, F. A., Green, D. H. & Roy, S. D. *J. Petrol.* **19**, 463–513 (1978).
37. Frey, F. A. & Green, D. H. *Geochim. cosmochim. Acta* **38**, 1023–1059 (1974).
38. Sclater, J. G., Jaupart, C. & Galson, D. *Rev. Geophys. Space Phys.* **18**, 269–312 (1980).
39. Jordan, T. H. *Phil. Trans. R. Soc.* (in the press).
40. Jacobsen, S. B. & Wasserburg, G. J. *J. geophys. Res.* **84**, 7429–7445 (1979).
41. Weaver, B. L. & Tarney, J. *Nature* **286**, 342–346 (1980).
42. Wise, D. U., in *The Geology of Continental Margins* (eds Burke, C. A. & Drake, C. L.) 45–58 (Springer, New York, 1974).

43. Armstrong, R. L. *Rev. Geophys. Space Phys.* **6**, 175–200 (1968).
44. Hess, H. H. in *Petrologic Studies* (eds Engel, A. E. J., James, H. L. & Leonard, B. F.) 599–620 (Geological Society of America, New York, 1962).
45. DePaolo, D. P. *Geochim. cosmochim. Acta* **44**, 1185–1196 (1980).
46. DePaolo, D. J. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **43**, 615–627 (1979).
47. Oxburgh, E. R. & Parmentier, E. M. *Phil. Trans. R. Soc. A288*, 415–428 (1978).
48. McKenzie, D. P. *Geophys. J. R. astr. Soc.* **58**, 689–715 (1979).
49. Richter, F. M. & Ribe, N. M. *Earth planet. Sci. Lett.* **43**, 212–222 (1979).
50. O'Connell, R. J. & Hager, B. H. *EOS* **61**, 373 (1980).
51. Davies, G. F. *Earth planet. Sci. Lett.* **44**, 231–238 (1979).
52. Davies, G. F. *J. geophys. Res.* **85**, 7108–7114 (1980).
53. DePaolo, D. J. *Earth planet. Sci. Lett.* **43**, 201–211 (1979).

Structure of the neuraminidase gene in human influenza virus A/PR/8/34

Stan Fields & Greg Winter

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

George G. Brownlee

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

The complete structure of the neuraminidase gene in influenza A/PR/8/34 has been determined by cloning into the bacteriophage M13 and sequencing with dideoxynucleotide chain terminators. The gene is 1,413 nucleotides long, codes for a protein of 454 amino acids and has five potential glycosylation sites. We suggest that the neuraminidase, unlike the influenza haemagglutinin, is oriented with its N-terminus buried in the viral membrane.

INFLUENZA A viruses have two surface antigens, haemagglutinin and neuraminidase, both of which can undergo extensive antigenic variation. The haemagglutinin binds the virus to sialic acid-containing receptors on the cell, but the role of the neuraminidase is still not fully understood (for reviews see refs 1, 2). The enzyme cleaves terminal *N*-acetylneuraminic acid (sialic acid) from carbohydrate chains in glycoproteins³, and its function may be to remove the carbohydrate residues from the virus envelope itself, thus eliminating receptors for the haemagglutinin and preventing self-aggregation of virion particles⁴. The removal of sialic acid also seems to facilitate proteolytic cleavage of the haemagglutinin, which is essential for infectivity^{5,6}, by exposing the site of cleavage on the haemagglutinin^{7,8}.

Although the haemagglutinin is the more significant factor in the immune response of the host organism, the neuraminidase is also involved in this response^{9,10}, as anti-neuraminidase antibodies can inhibit viral replication. Like the haemagglutinin, the neuraminidase can undergo either small changes in antigenicity^{11,12} (antigenic drift) or major changes (antigenic shift), in which the neuraminidases do not cross-react serologically¹¹. Antigenic shift in human influenza viruses occurred in 1957, when the N2 subtype replaced the N1. A comparison of the N-terminal amino acid sequences predicted from RNA sequencing¹³ showed that within a subtype there is a high degree of amino acid conservation, but between the different subtypes there is little homology after the first 12 amino acids.

Using recombinant DNA methods, we have sequenced the neuraminidase of the human strain A/PR/8/34 (H1N1, formerly designated H0N1) to provide a basic framework for both understanding the function of the protein and to further study the mechanism of antigenic drift and shift. From the sequence has emerged the possibility that in contrast to the haemagglutinin^{14–20}, the neuraminidase is bound to the viral membrane by its N-terminus.

Cloning and sequence analysis in M13

Three separate strategies were used to determine the sequence. As part of a 'shotgun' approach²¹ to sequence the eight single-stranded RNA segments in the influenza genome^{22–24}, restriction fragments derived from the whole genome were cloned into

the bacteriophage M13mp2 (ref. 25) or one of its derivatives^{26,27} and sequenced by the dideoxy technique²⁸. One *Hae*III clone contained the sequence of 444 nucleotides at the 3' end of the neuraminidase virion RNA (vRNA) as seen from a match with limited RNA sequence data for the related fowl plague virus (FPV) neuraminidase gene^{29,30}. Two other clones, derived from

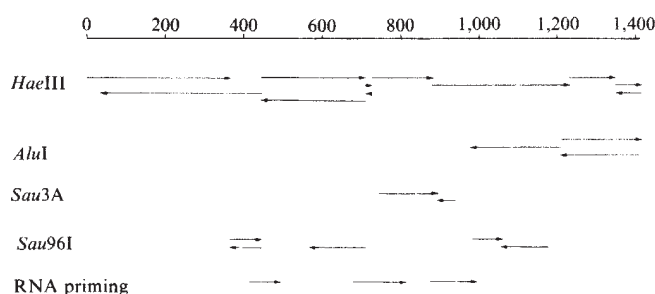


Fig. 1 Summary of sequence evidence. Virion RNA was polyadenylated and double-stranded cDNA prepared using oligo(dT)_{12–18} priming and then loopback of the 3' end of the cDNA as described elsewhere²¹. Alternatively, the band 6 cDNA derived from 20 µg RNA was eluted from a 2.8% acrylamide/7 M urea gel³⁴ and annealed in 40 µl of 0.05 M NaCl, 0.01 M Tris-HCl pH 7.4, 0.01 M MgCl₂, 1 mM dithiothreitol (DTT) to 80 pmol of a 13-long synthetic oligodeoxynucleotides primer (see text) by boiling and then slowly cooling. Deoxynucleotide triphosphates to 0.5 mM and 2U DNA polymerase (Klenow fragment, Boehringer) were added and the reaction incubated for 1 h at 20 °C. Double-stranded molecules were cut by restriction enzymes; *Hae*III and *Alu*I fragments were blunt-end ligated to 10 ng *Hind*II-cut M13mp7 (ref. 27) in 10 µl of 0.05 M Tris-HCl pH 7.4, 0.01 M MgCl₂, 0.01 M DTT, 1mM rATP and 0.4U T4 DNA ligase. *Sau*96I fragments were first blunt-ended by incubation with the Klenow fragment of DNA polymerase³⁶ and then ligated as above. *Sau*3A fragments were ligated as described elsewhere²¹. DNA sequencing of M13 clones used a 30-base pair cloned primer⁶⁵ or 17-long oligodeoxynucleotide synthetic primer (from Dr M. J. Gait). RNA priming with restriction fragment primers was as described elsewhere²¹. Sequence reactions were electrophoresed on 6% acrylamide/7 M urea thin gels⁶⁶.

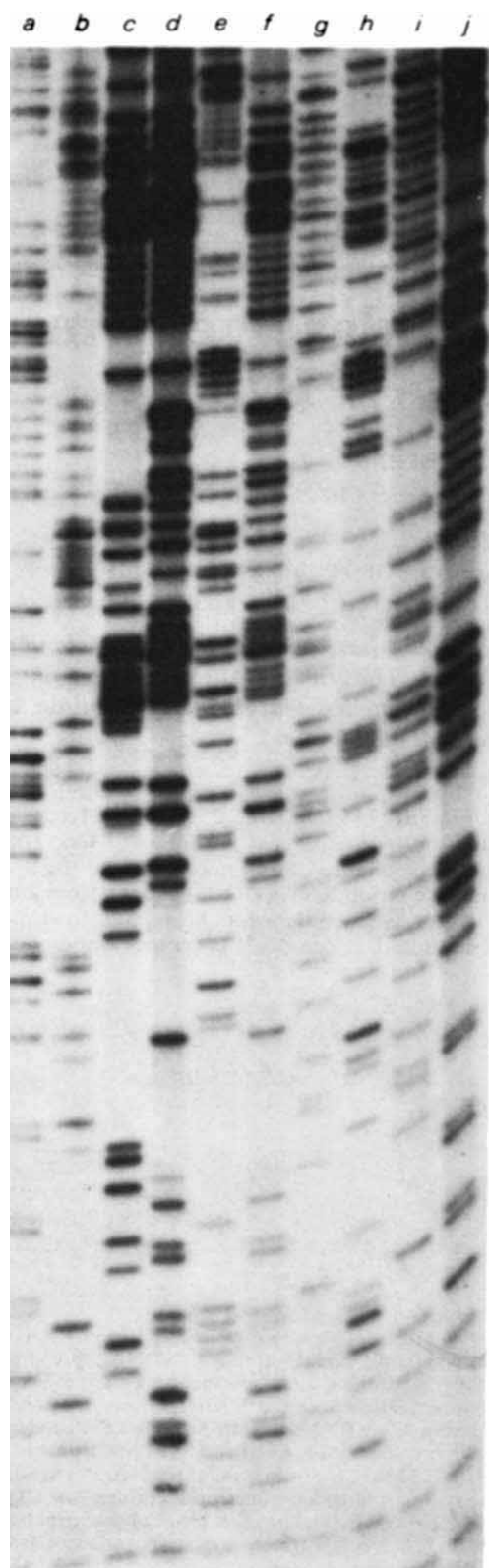


Fig. 2 Screening clones by dideoxythymidine tracks. *Hae*III clones were isolated as described in Fig. 1 legend. Screening using dideoxythymidine patterns as characterization³⁶ was on a 6% acrylamide/7 M urea thin gel⁶⁶ electrophoresed at 30 mA for 1.5 h. The 10 clones screened on this gel correspond to the following sequences (see Fig. 3): *a*, positions 1,347 to 1,413; *b*, 1,346–1,233; *c*, 1,346–1,233; *d*, 1,233–1,346; *e*, 727–878; *f*, 1,233–1,346; *g*, 710–726; *h*, 879–1,232; *i*, 710–726; and *j*, 1,347–1,413. Therefore clones *a* and *j* are identical, as are *g* and *i*. Clones *b* and *c* contain the same *Hae*III fragment but the opposite strand to clones *d* and *f*.

Table 1 Amino acid composition of the A/PR/8/34 neuraminidase

Amino acid	No.	Amino acid	No.
Ala (A)	16	Leu (L)	21
Arg (R)	20	Lys (K)	23
Asn (N)	25	Met (M)	7
Asp (D)	25	Phe (F)	16
Cys (C)	19	Pro (P)	21
Glu (E)	17	Ser (S)	51
Gln (Q)	11	Thr (T)	29
Gly (G)	44	Trp (W)	16
His (H)	10	Tyr (Y)	14
Ile (I)	41	Val (V)	28
Total no. amino acids		454	

*Hae*III and *Alu*I shotgun digests, contained an additional 500 nucleotides of neuraminidase sequence when overlapped by the other approaches.

To extend the 3' sequence, we prepared double-stranded cDNA (dsDNA) specific for segment 6, the neuraminidase gene^{31–33}, after first purifying the single-stranded cDNA (ssDNA) on an acrylamide gel³⁴. Clones derived from restriction digests of this dsDNA (Fig. 1) provided all of the remaining sequence of the neuraminidase gene, but not every overlap. The *in vitro* back-copy reaction was greatly facilitated by the use of a 13-long oligodeoxynucleotide primer³⁵ complementary to the 3' termini of the ssDNAs²⁹, which allowed synthesis of the complete sequence corresponding to the 5' end of the vRNA.

The final three overlaps were obtained from the third approach, in which restriction fragments derived from the replicative form of selected recombinant M13 clones were used as primers for dideoxy sequencing on a total vRNA template²¹.

Our approach was such that most sequences were determined more than once. As clones were obtained in either orientation in M13, 93% of the neuraminidase gene was determined on both strands, and of the remainder, most was sequenced from at least two independent clones. A total of 25 clones and three RNA primings was used in the complete sequence determination.

Because no clone is identified before its sequencing, particular restriction digests may yield an individual fragment in the same orientation in many clones. This problem of redundant information, inherent in a 'shotgun' strategy, was avoided by eliminating excess clones with a rapid characterization³⁶ that used only one of the four dideoxy tracks of a normal sequence reaction. Figure 2 shows the pattern of 'dideoxythymidine tracks' for one set of *Hae*III fragments derived from band 6 dsDNA. Such tracks show whether a clone has been obtained before and can indicate the length of an insert by inspection of the pattern for the M13 bands which follow the insert.

Nucleotide sequence of the neuraminidase gene

Figure 3 shows the complete sequence of the neuraminidase gene as the complementary RNA, corresponding to the messenger sense. The sequence is 1,413 nucleotides long, including 20 nucleotides 5' to the first AUG and 31 nucleotides 3' to the last codon. (The actual structure of the mRNA produced by the virus has the cap and several nucleotides donated by a host mRNA³⁷ and is an incomplete transcript of the vRNA³⁸.) Only one reading frame is open, beginning with the AUG at position 21 and terminating at a UAG at position 1,383; the next in-frame AUG is at position 450 and could code for a protein with a molecular weight (MW) of only 34,000 (see below). The other two reading frames are blocked by numerous termination codons. The predicted protein is co-linear with the RNA and occupies 97% of the coding capacity of the gene. Whereas band 8 (refs 39, 40) and possibly band 77 (ref. 21) of influenza virus code for two proteins, the neuraminidase gene is similar to the haemagglutinin gene^{16–20} in that it codes for a single protein. Previous sequencing of the A/PR/8/34 neuraminidase gene has determined limited sequences at the 3'

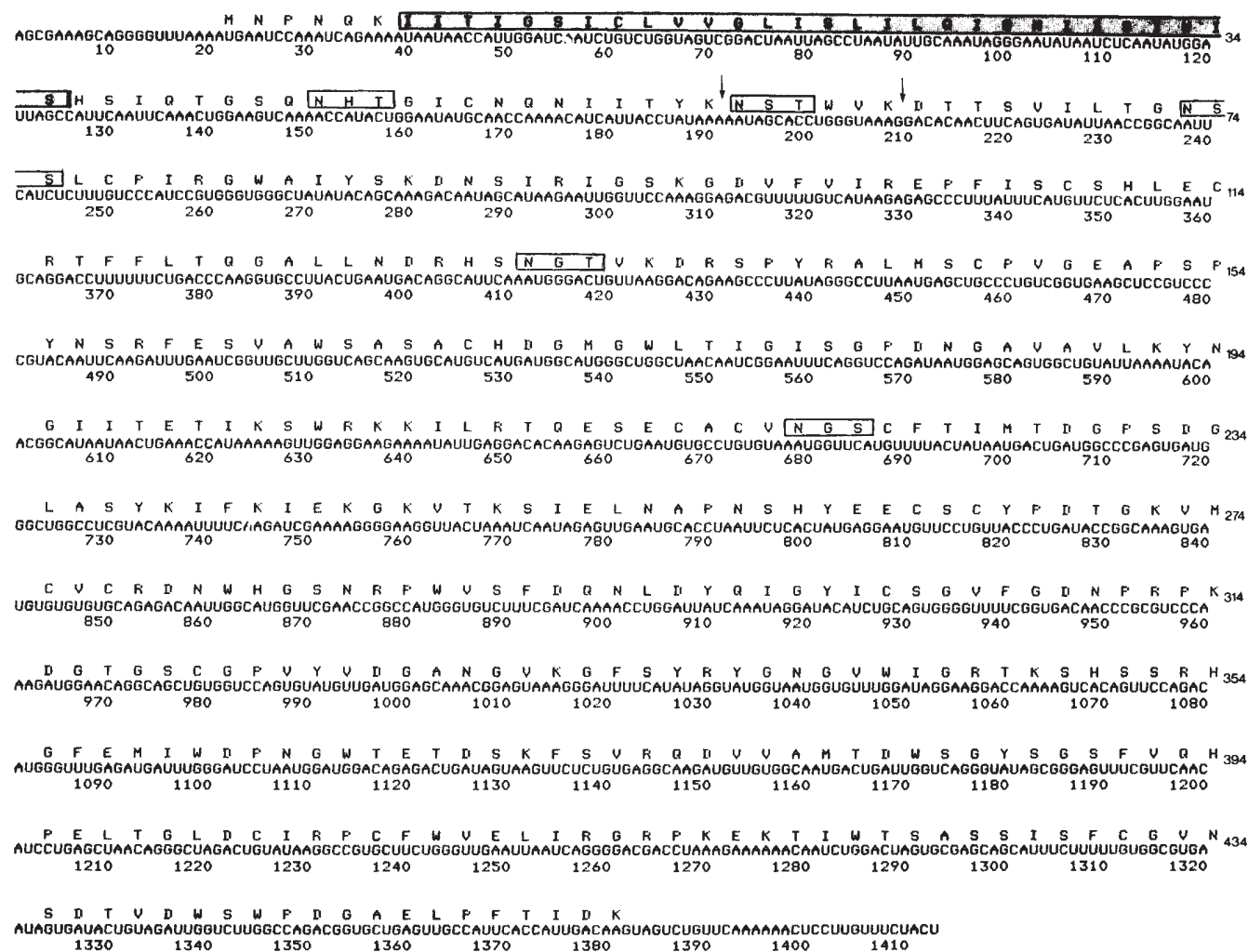


Fig. 3 Sequence of the neuraminidase gene in complementary RNA sense. The protein predicted from the sequence is given in the one-letter code (see Table 1); amino acid residues are numbered at the far right and potential glycosylation sites are boxed. The N-terminal hydrophobic region is shaded, and arrows indicate two possible trypsin cleavage points after this region (see text).

(ref. 41) and 5' (ref. 42) termini of the vRNA. Our sequence differs from these results by only one nucleotide at the 3' end of the vRNA.

The considerable antigenic variation of the haemagglutinin and neuraminidase might be reflected in sequence heterogeneity within a strain. Thus a single cloned copy could be derived from an RNA segment which is a minor variant in the population. The M13 shotgun strategy avoids this problem because an average sequence is obtained, reflecting many individual RNA segments. We found remarkably little variation in the neuraminidase gene, with only two positions showing nucleotide differences (position 66 was an A residue and 769 a G residue, each resulting in a conservative amino acid change). However, in both cases the sequences from more than one clone were in agreement—only that from a single clone differed.

Structure of the neuraminidase

The sequence of the neuraminidase gene codes for a protein of 454 amino acids (see Table 1), MW 50,087, excluding carbohydrate. The neuraminidase exists in the viral membrane as a tetramer⁴³; our predicted molecular weight agrees with previous estimates of the subunit molecular weight which ranged from 48,000 to 63,000 (refs 43–45) including 20% carbohydrate⁴³. An amino acid composition of the neuraminidase in A/Bel/42 (ref. 46), also an N1 strain, correlates reasonably well with the composition of the A/PR/8/34 protein.

Potential carbohydrate attachment sites Asn-X-Ser or Asn-X-Thr⁴⁷ occur at amino acid positions 44, 58, 73, 121 and 220 (Fig. 3). In the haemagglutinin (H2N2 A/Japan/305/57), five out of six potential sites are glycosylated²⁰, and only the normally unglycosylated Asn-Pro-Ser sequence⁴⁷ is not. Thus it is likely that most or all of the neuraminidase potential sites, none of which contains proline, has carbohydrate attached by N-glycosidic linkage. The amino acid composition (Table 1) shows that the neuraminidase is rich in cysteine (19 residues), and although evidence suggests that disulphide bonds link monomers to form a dimer⁴⁵, the positions or number of the cysteine bridges are not known.

A Chou-Fasman⁴⁸ prediction of the secondary structure of the protein indicates that the molecule is mainly β -sheet, with helical potential at amino acid residues 109(Cys)–129(His), 201(Ile)–219(Val) and 235(Leu)–249(Thr). A helix wheel plot⁴⁹ confirms that the second of these three regions is likely to be an α helix. Proline residues (21 in all) are scattered throughout the molecule, especially at predicted regions of β bend, in contrast to the HA₂ subunit of the haemagglutinin which contains few or no proline residues^{14–20} and is thought to be highly helical¹⁴.

Possible membrane attachment site

The influenza virus neuraminidase is an integral membrane protein and can only be released from the viral envelope by detergent or protease treatment⁵⁰. Previously it was suggested⁴⁵



Fig. 4 Comparison of membrane attachment sites. The N-terminal hydrophobic sequence of the neuraminidase is compared with the membrane attachment regions of isomaltase⁵⁹, glycophorin⁵⁴, vesicular stomatitis virus (VSV) glycoprotein⁵⁷, membrane-bound IgM⁵⁶ and the influenza virus haemagglutinin of A/PR/8/34 (unpublished data). The boxed region indicates the uncharged residues thought to span the membrane. The sequences have been aligned to show amino acid homology with the neuraminidase.

that a hydrophobic region was present at either the N- or C-terminus because the trypsin-released enzyme is smaller (by an estimated MW of 7,000 (ref. 45) or 12,000 (ref. 43)) than that derived from SDS disruption of the virion. This hydrophobic region would anchor the neuraminidase to the lipid bilayer of the virus. An examination of the A/PR/8/34 neuraminidase (Fig. 3) shows only one major hydrophobic region, located near the N-terminus and extending from amino acid residue 7 to 35. Of these 29 residues, 18 are hydrophobic, 11 are neutral and none is charged. The first trypsin-sensitive residues following this hydrophobic region are at positions 57(Lys) and 63 (Lys), and tryptic cleavage at either of these positions would result in a loss of MW of ~6,000, which is largely consistent with protein chemistry studies^{43,45}. A stretch of 16 uncharged amino acid residues (420–435) lies near the C-terminus, but of these only 5 are hydrophobic and seem unlikely to be sufficient for attachment of an integral membrane protein. Additional evidence favouring an N-terminal attachment comes from the finding that the small fragment lost on protease isolation of the enzyme from B/Lee⁴⁵ or FVPV⁵¹ is rich in carbohydrate. In the A/PR/8/34 neuraminidase, potential carbohydrate attachment sites near the N-terminus occur at residues 44 and 58, and at least the first of these sites would be contained within the tryptic peptide presumed to be lost. No attachment sites lie within the C-terminal half of the molecule.

This membrane orientation would contrast with that of the influenza virus haemagglutinin⁵², and also of such membrane glycoproteins as erythrocyte glycophorin^{53,54}, HLA antigens⁵⁵, membrane-bound immunoglobulin μ chains⁵⁶ and vesicular stomatitis virus glycoprotein⁵⁷, all of which have C-terminal hydrophobic sequences. However, attachment of proteins to membranes through a hydrophobic N-terminus does have a precedent in the orientation of intestinal brush border aminopeptidase⁵⁸ and isomaltase^{58–60}. Figure 4 shows a comparison of the N-terminal hydrophobic sequence of the neuraminidase with other sequences involved in spanning the lipid bilayer; glycophorin shows the greatest sequence homology to the neuraminidase but all the sequences are similarly hydrophobic.

There is no experimental evidence to indicate whether the N-terminal hydrophobic sequence of the neuraminidase is a 'signal sequence' which is cleaved off⁶¹, or whether it remains

attached to the membrane, as has been suggested for the N-terminal region of isomaltase⁵⁹. However, in the two human subtypes, the total conservation of the first 12 amino acid residues¹³, of which the first six are hydrophilic, suggests that these have some critical role, which might be to interact with another protein or neuraminidase subunit. (Signal sequences of the haemagglutinins, which are cleaved off⁶², show hydrophobicity²⁰ but no conservation of amino acid sequence.) We suggest that the neuraminidase, like isomaltase, has an 'extended signal sequence'⁵⁹ which both transfers the protein across the membrane and remains in the bilayer to anchor the protein, with the first six residues protruding into the virion as a hydrophilic tail. The signal sequence may not be cleaved because the recognition sequence for the signal peptidase is lacking or unexposed⁵⁹. Such an N-terminal attachment favours a loop model⁶³ of protein translocation across a membrane, in which the signal sequence inserts into the lipid bilayer as a loop, with the first few amino acid residues remaining on the cytoplasmic side of the membrane. This hypothesis concerning N-terminal attachment of the neuraminidase could be tested directly by amino acid sequence analysis of the intact and trypsin-released enzyme.

Concluding remarks

The sequence of the neuraminidase gene allows us to suggest a possible role for a long hydrophobic sequence at the protein's N-terminus, but the regions involved in enzymatic activity and antigenicity remain undefined. Future sequences of other neuraminidase genes should reveal regions of homology within and between subtypes, and thus elucidate the phenomena of antigenic drift and shift at the molecular level. A more complete understanding of the neuraminidase, however, must await a three-dimensional structure of the molecule, which is at the preliminary stage for the N2 subtype⁶⁴.

We thank Mrs M. A. Robertson for supplying virus and RNA, R. Henderson, M. Bretscher and D. L. Bentley for discussions, A. D. McLachlan for the Chou-Fasman prediction and J. Messing for M13 vectors. S.F. was supported by a NSF predoctoral fellowship and G.W. by a research fellowship from Trinity College, Cambridge.

Received 7 November 1980; accepted 15 January 1981.

- Schulze, I. T. in *The Influenza Viruses and Influenza* (ed. Kilbourne, E. D.) 53–82 (Academic, London, 1975).
- Bucher, D. & Palese, P. in *The Influenza Viruses and Influenza* (ed. Kilbourne, E. D.) 83–123 (Academic, London, 1975).
- Gottschalk, A. *Biochim. biophys. Acta* **23**, 645–646 (1957).
- Palese, P. *et al. Virology* **61**, 397–410 (1974).
- Klenk, H.-D. *et al. Virology* **68**, 426–439 (1975).
- Lazarowitz, S. G. & Choppen, P. W. *Virology* **68**, 440–454 (1975).
- Schulman, J. L. & Palese, P. *J. Virol.* **24**, 170–176 (1977).
- Nakajima, S. & Sugiyama, A. *Virology* **101**, 450–457 (1980).
- Schulman, J. L., Khakpour, M. & Kilbourne, E. D. *J. Virol.* **2**, 778–786 (1968).
- Rott, R., Becht, H. & Orlich, M. *J. gen. Virol.* **22**, 35–41 (1974).

- Paniker, C. K. *J. J. gen. Virol.* **2**, 385–394 (1968).
- Schulman, J. L. & Kilbourne, E. D. *Proc. natn. Acad. Sci. U.S.A.* **63**, 326–333 (1969).
- Blok, J. & Air, G. M. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 125–133 (Elsevier, Amsterdam, 1980).
- Ward, C. W. & Dopheide, T. A. *Br. med. Bull.* **35**, 51–56 (1979).
- Waterfield, M. D. *et al. Br. med. Bull.* **35**, 57–63 (1979).
- Porter, A. G. *et al. Nature* **282**, 471–477 (1979).
- Min Jou, W. *et al. Cell* **19**, 683–696 (1980).
- Both, G. W. & Sleight, M. J. *Nucleic Acids Res.* **8**, 2561–2575 (1980).
- Verhoeyen, M. *et al. Nature* **286**, 771–776 (1980).
- Gething, M.-J. *et al. Nature* **287**, 301–306 (1980).
- Winter, G. & Fields, S. *Nucleic Acids Res.* **8**, 1965–1974 (1980).
- McGeoch, D. J., Fellner, P. & Newton, C. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3045–3049 (1976).

23. Palese, P. & Schulman, J. L. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2142–2146 (1976).
24. Scholtissek, C. *et al. Virology* **74**, 332–344 (1976).
25. Gronenborn, B. & Messing, J. *Nature* **272**, 375–377 (1978).
26. Rothstein, R. J. *et al. Meth. Enzym.* **68**, 101–110 (1979).
27. Messing, J., Crea, R. & Seeburg, P. H. *Nucleic Acids Res.* **9**, 309–321 (1981).
28. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
29. Skehel, J. J. & Hay, A. J. *Nucleic Acids Res.* **5**, 1207–1219 (1978).
30. Robertson, J. S. *Nucleic Acids Res.* **6**, 3745–3757 (1979).
31. Palese, P. & Schulman, J. L. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2142–2146 (1976).
32. Scholtissek, C. *et al. Virology* **74**, 332–344 (1976).
33. Inglis, S. C., McGeoch, D. J. & Mahy, B. W. J. *Virology* **78**, 522–536 (1977).
34. Sleight, M. J., Both, G. W. & Brownlee, G. G. *Nucleic Acids Res.* **6**, 1309–1321 (1979).
35. Winter, G., Fields, S., Gait, M. J. & Brownlee, G. G. *Nucleic Acids Res.* **9**, 237–245 (1981).
36. Sanger, F. *et al. J. molec. Biol.* **143**, 161–178 (1980).
37. Krug, R. M., Broni, B. A. & Bouloy, M. *Cell* **18**, 329–334 (1979).
38. Hay, A. J. *et al. Nucleic Acids Res.* **4**, 4197–4209 (1977).
39. Inglis, S. C. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 3790–3794 (1979).
40. Lamb, R. A. & Choppin, P. W. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4908–4912 (1979).
41. Desselberger, U., Zamecnik, P. & Palese, P. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 169–179 (Elsevier, Amsterdam, 1980).
42. Desselberger, U. *et al. Gene* **8**, 315–328 (1980).
43. Wrigley, N. G. *et al. Virology* **51**, 525–529 (1973).
44. Gregoriades, A. *Virology* **49**, 333–336 (1972).
45. Lazdins, I., Haslam, E. A. & White, D. O. *Virology* **49**, 758–765 (1972).
46. Laver, W. G. & Baker, N. J. *gen. Virol.* **17**, 61–67 (1972).
47. Neuberger, A. *et al. in The Glycoproteins: Their Composition, Structure and Function* (ed. Gottschalk, A.) 450–490 (Elsevier, Amsterdam, 1972).
48. Chou, P. Y. & Fasman, G. D. *Biochemistry* **13**, 211–245 (1974).
49. Schiffer, M. & Edmundson, A. B. *Biophys. J.* **7**, 121–135 (1967).
50. Drzeniek, R., Seto, J. & Rott, R. *Biochim. biophys. Acta* **128**, 547–558 (1966).
51. Rott, R. *et al. Naturforsch. B.* **27**, 227–233 (1972).
52. Skehel, J. J. & Waterfield, M. D. *Proc. natn. Acad. Sci. U.S.A.* **72**, 93–97 (1975).
53. Tomita, M. & Marchesi, V. T. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2964–2968 (1975).
54. Furthmayr, H. *et al. Archs Biochem. Biophys.* **185**, 21–29 (1978).
55. Robb, R. J., Terhorst, C. & Strominger, J. L. *J. biol. Chem.* **253**, 5319–5324 (1978).
56. Rogers, J. *et al. Cell* **20**, 303–312 (1980).
57. Rose, J. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3884–3888 (1980).
58. Maroux, S. & Louvard, D. *Biochim. biophys. Acta* **419**, 189–195 (1976).
59. Frank, G. *et al. FEBS Lett.* **96**, 183–188 (1978).
60. Brunner, J. *et al. J. biol. Chem.* **254**, 1821–1828 (1979).
61. Blobel, G. & Dobberstein, B. *J. Cell Biol.* **67**, 835–851 (1975).
62. Elder, K. T. *et al. Virology* **95**, 343–350 (1979).
63. Inouye, M. & Halegoua, S. *CRC crit. Rev. Biochem.* **7**, 339–371 (1980).
64. Colman, P. M., Tulloch, P. A. & Laver, W. G. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 351–356 (Elsevier, Amsterdam, 1980).
65. Anderson, S. *et al. Nucleic Acids Res.* **8**, 1731–1743 (1980).
66. Sanger, F. & Coulson, A. R. *FEBS Lett.* **87**, 107–110 (1978).

A plasmid DNA primase active in discontinuous bacterial DNA replication

Brian M. Wilkins & Graham J. Boulnois

Department of Genetics, University of Leicester, Leicester LE1 7RH, UK

Erich Lanka

Max-Planck-Institut für Molekulare Genetik, Abteilung Schuster, Berlin-Dahlem, FRG

A DNA primase encoded by an IncIα plasmid promotes efficient DNA replication in a primase-defective mutant of Escherichia coli. This finding implies that the plasmid enzyme can prime discontinuous DNA synthesis of the bacterial chromosome. The plasmid gene encodes two large, antigenically related proteins which differ from E. coli primase.

A CENTRAL feature of DNA replication is the discontinuous synthesis of the strand that grows in the overall direction of 3' to 5'. DNA polymerases cannot initiate synthesis of the fragments involved in the process because these enzymes require the 3'-OH terminus of an oligonucleotide primer (for review see ref. 1). During replication of the *Escherichia coli* chromosome, the primer sequences are synthesized by the product of the *dnaG* gene², called primase³. This 60,000 molecular weight (M_r) single subunit protein⁴ is thought to act in a multi-enzyme complex that includes *dnaB* product⁵. Although certain virulent bacteriophages are known to encode DNA primases, these enzymes act in phage DNA replication in cooperation with specific phage DNA polymerases^{6–9}.

Conjugative plasmids of the *Iα* and *Iγ* incompatibility groups¹⁰, represented by ColI, R64, R144 and R621a, can suppress temperature-sensitive *dnaG* mutations, allowing DNA replication and colony formation at high temperature^{11,12}. The relevant gene is designated *sog*, conferring a phenotype of *Sog*⁺. Suppression of *dnaG* lesions requires derepression of the plasmid transfer genes by a *drd* mutation, implying that the physiological role of *sog* product may be in conjugation. Recently, a novel DNA primase has been purified from bacteria containing the IncIα plasmid R64*drd-11* and characterized¹⁴. The enzyme has a low template specificity *in vitro*; it can substitute for the activities of *dnaB-dnaC-dnaG* products, *E. coli* RNA polymerase and *dnaG* product in priming synthesis of DNA complementary to viral strands of phages ΦX174, fd and G4, respectively.

The studies reported here show that the IncIα plasmid product responsible for suppressing bacterial *dnaG* mutations is a DNA primase. This finding implies that the plasmid enzyme can substitute for *E. coli* primase to promote efficient discontinuous synthesis of the bacterial chromosome and indicates that primer-generating enzymes other than *dnaG* protein can act in bacterial DNA replication. The approach has involved the

cloning of the *sog* gene of ColI*drd-1* and analysis of proteins encoded by the parental and recombinant plasmids. The *sog* gene is shown to encode two antigenically related proteins. Primase activity is localized in part of the larger molecule.

Table 1 Plasmids that suppress *dnaG3* specify the *Iα* plasmid primase

Strain	Colony formation 40 °C/30 °C	DNA synthesis at 41 °C (10 ⁻³ c.p.m. per ml per A ₄₅₀ unit)	Primase activity (mU per mg CCE)
BW86	<10 ⁻⁷	11	ND
BW86 (pBR325)	<10 ⁻⁷	16	ND
BW86 (ColI)	2.7 × 10 ⁻⁷	22	1
BW86 (ColI <i>drd-1</i>)	2.1 × 10 ⁻¹	342	213
BW86 (pLG215)	1.8 × 10 ⁻⁴	262	182
BW86 (pLG214)	1.0	341	221

Colony formation gives the fractional yield of colonies at the indicated temperatures. Overnight cultures in nutrient broth were diluted suitably and plated on prewarmed nutrient agar. Thymine (20 μg ml⁻¹) was added to all media. DNA synthesis indicates ³H-thymine incorporated in 60 min at 41 °C. Bacteria, growing exponentially in SGC medium containing deoxyguanosine and thymine (3 μg ml⁻¹)¹¹, were incubated at 41 °C for 5 min before addition of ³H-thymine to 7.5 μCi ml⁻¹ (320 mCi mmol⁻¹). A₄₅₀ of cultures was measured at zero time. Primase was assayed using bacteria growing exponentially in broth. 2 × 10⁹ bacteria were gently lysed using the non-ionic detergent Brij 58 in conjunction with lysozyme²⁶. Following centrifugation (80,000 g, 30 min), primase activity and protein concentration were measured as described previously¹⁴. All liquid media for strains harbouring pBR325, pLG214 or pLG215 contained tetracycline (10 μg ml⁻¹) and agar contained ampicillin (25 μg ml⁻¹). 1 unit of primase incorporated 1 μmol of dTMP in the conditions used. CCE, crude cell extract; ND, not detectable.

Cloning of the *sog* gene

The small multicopy plasmid pBR325 was used as the vector for cloning. This plasmid¹⁵ carries genes conferring resistance to ampicillin (Ap^r), tetracycline (Tc^r) and chloramphenicol (Cm^r) and contains a single endonuclease *Eco*RI cleavage site in the latter. Fragments of *ColI**drd-1*, generated by *Eco*RI restriction endonuclease, were ligated into this site in pBR325 and the DNA was used to transform a temperature-sensitive *dnaG3* mutant (BW86, *dnaG3 leu thyA deoB rpsL Δ chlA-uvrB ColI*; ref. 13). Transformants harbouring *Sog*⁺ plasmids could be isolated using the following strategy. Ap^r transformants were first selected at 30 °C and allowed to grow into small colonies, when they were replica plated. About 1% of the replicas grew at 40 °C to give distinct colonies; these were found to contain *Sog*⁺ recombinant plasmids.

The plasmid (pLG211) in one of these strains was characterized. Cleavage with *Eco*RI restriction endonuclease generated three DNA fragments with estimated sizes of 5.9 kilobases (which corresponds to the vector), 7.9 and 0.27 kilobases. The smallest of these DNA fragments was next removed by cleaving pLG211 with *Eco*RI endonuclease followed by ligation to give pLG215. The latter plasmid specifies resistance to ampicillin

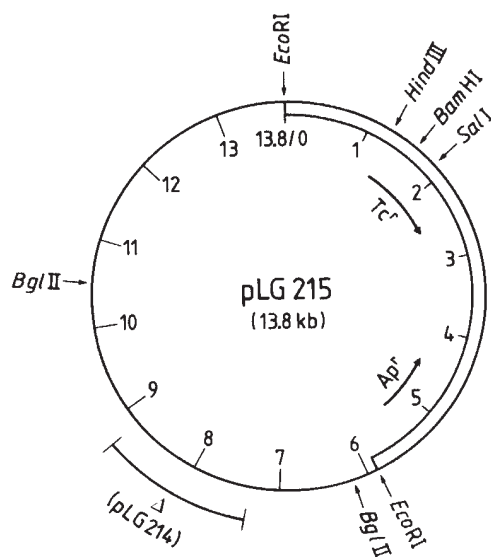


Fig. 1 Map of pLG215 showing the position of the 1.5-kilobase deletion mutation present in pLG214. The vector, pBR325, is drawn in double lines (0 to 5.9 kilobases). The positions of the Tc^r and Ap^r genes are derived from maps of pBR322 and pBR325 (refs 15, 27). Mapping was carried out by subjecting restriction endonuclease-cleaved DNA to electrophoresis in 0.5% or 2% agarose gels using procedures similar to those described elsewhere^{28,29}, except that double digestions were carried out in 6.6 mM MgCl_2 , 60 mM NaCl, 6.6 mM 2-mercaptoethanol and 6.6 mM Tris-HCl (pH 7.4). Molecular weight markers included phage λ DNA cleaved with *Eco*RI and *Hind*III endonucleases in single or double digests³⁰ and *Taq*YI-cleaved pBR322 (ref. 31). Lengths (kilobases) of DNA fragments in *Bgl*II + *Eco*RI double digests were 5.9, 4.4, 3.3 and 0.23. Lengths of DNA fragments generated by *Bgl*II + *Hind*III endonucleases were 4.9, 4.6 and 4.4 kilobases. The 4.4-kilobase DNA fragment was replaced by a 2.9-kilobase fragment in corresponding digests of pLG214. The *Bam*HI and *Sal*I cleavage sites were mapped using sizes of fragments generated in appropriate double digests with *Hind*III or *Eco*RI endonucleases. Construction of pLG215 is summarized in the text. pBR325 was isolated by a cleared lysate method³². *ColI**drd-1* was purified using a Sarkosyl lysis method³³ with the modification that DNA in supernatant was concentrated by ethanol precipitation following extraction with phenol. Ligations were done at 4 °C in 60 mM Tris-HCl (pH 7.4), 10 mM MgCl_2 , 10 mM dithiothreitol, 0.1 mM ATP, 0.005% gelatin and T4 DNA ligase (from Dr Ron Wilson).

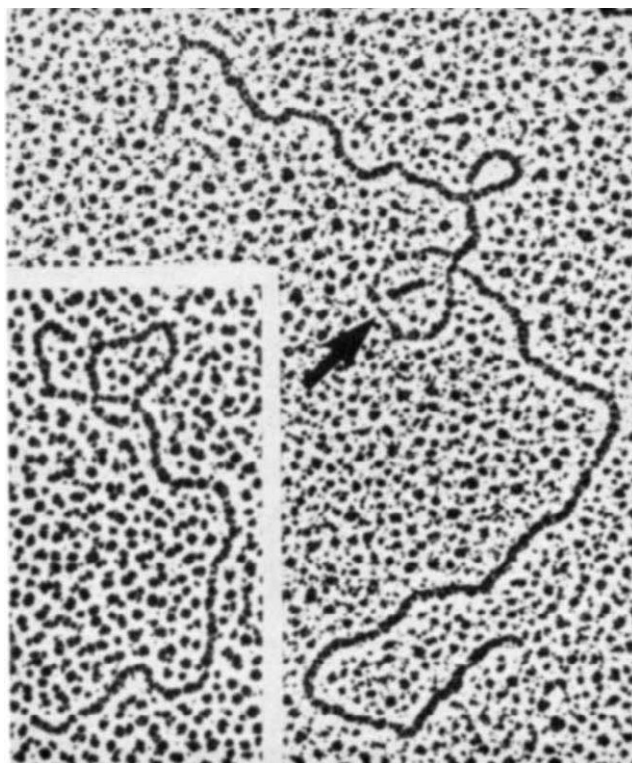


Fig. 2 Electron micrograph of a pLG214/pLG215 heteroduplex DNA molecule showing a deletion loop, indicated by the arrow. The plasmids were cleaved at the unique *Hind*III site (Fig. 1) before formation of heteroduplexes. The inset shows a heteroduplex DNA molecule prepared following digestion of pLG214 and pLG215 with *Eco*RI endonuclease. This digestion separates the cloned fragment of *ColI**drd-1* from the vector (Fig. 1). The mean fractional length of the short arm of 18 molecules was 0.22 (s.d. $\pm$ 3.7%) of the duplex DNA. The estimated length of the duplex DNA is 6.4 kilobases, which is the size of the relevant *Eco*RI fragment in pLG214. Molecules were prepared and examined using methods described elsewhere³⁴.

and to tetracycline but not to chloramphenicol, indicating insertion-inactivation of the Cm^r gene of pBR325. Plasmid pLG215 retains the 7.9-kilobase *Eco*RI restriction fragment of *ColI**drd-1* which was found to contain two endonuclease *Bgl*II cleavage sites (Fig. 1). Table 1 shows that the plasmid suppresses the *dnaG* mutation in BW86 as measured by extensive DNA synthesis at 41 °C and colony formation at 40 °C. However, by both criteria, pLG215 suppressed the *dnaG* lesion less efficiently than *ColI**drd-1*.

BW86(pLG215) is deficient in accumulating DNA at high temperature. This was established by monitoring absorbance and trichloroacetic acid (TCA)-precipitable ³H-thymine in a uniformly labelled culture: following transfer of an exponentially growing culture from 31 °C to 41 °C, A_{450} increased by a factor of 7.0 in 120 min at high temperature whereas DNA increased by a factor of 3.6. No such disparity was detected at 31 °C. This finding, together with the heterogeneous size of colonies formed at 40 °C, suggested that an undefined mutation might be necessary to allow BW86 (pLG215) to form colonies at high temperature. Chromosomal mutation is inferred from studies with six independently isolated colonies formed at 40 °C. The plasmids in these strains appeared identical to pLG215 as judged by analysis of *Eco*RI restriction digests and, following reintroduction of the plasmids into BW86, by the ability of the resulting transformants to form colonies at 40 °C. Plasmid-free segregants of the six original strains were temperature sensitive but they acquired a fully thermoresistant phenotype with high frequency in a second round of transformation with pLG215.

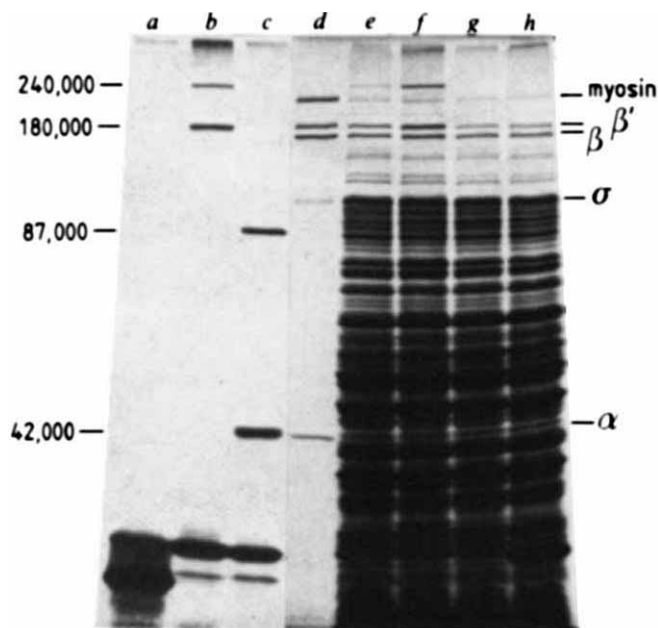


Fig. 3 Analysis by SDS-gel electrophoresis of proteins encoded by *Sog*⁺ plasmids. Lanes *a-c* show a fluorograph of ³⁵S-labelled proteins synthesized in minicells containing: *a*, pBR325; *b*, pLG215; *c*, pLG214. Lanes *e-h* show Coomassie blue-stained proteins in extracts of: *e*, BW86 (pLG215); *f*, BW84 (pLG216); *g*, BW86 (pLG214); *h*, BW86 (pBR325). Molecular weight standards (lane *d*) are myosin (*M_r* 212,000), *E. coli* RNA polymerase β' -subunit (*M_r* 165,000), β -subunit (*M_r* 155,000), σ -subunit (*M_r* 90,000) and α -subunit (*M_r* 39,000). Minicells were purified³⁵ from plasmid-containing derivatives of DS410 (ref. 36). Proteins were labelled for 90 min at 37 °C in M9 minimal medium containing ³⁵S-methionine (100 μ Ci ml⁻¹, 1 Ci mmol⁻¹) and Difco methionine assay medium (1.6 mg ml⁻¹). After a 10-min chase with methionine-supplemented broth, minicells were lysed by boiling in cracking buffer (62.5 mM Tris-HCl (pH 6.8), 3% SDS, 10% glycerol, 5% β -mercaptoethanol). About 150,000 c.p.m. were added per slot. Conditions for gel electrophoresis³⁷ and fluorography³⁸ have been described. Cell extracts were prepared from 1 ml of stationary phase bacteria at *A*₄₅₀ of 1. Cells were lysed in 30 μ l cracking buffer and the total lysate was added to the slot

Thus, the strains contain a chromosomal mutation which does not revert the *dnaG3* lesion but allows efficient colony formation at 40 °C in the presence of the *Sog*⁺ plasmid.

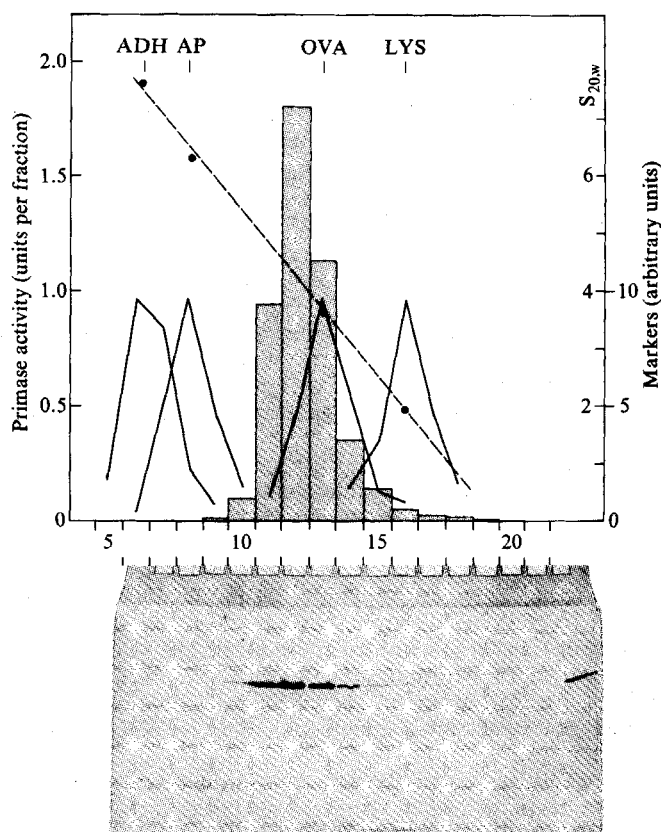
The recloning experiment which generated pLG215 from pLG211 gave rise to a BW86 transformant strain that was atypical in forming colonies at 40 °C with a frequency of 1. This phenotype was determined by a plasmid designated pLG214 (Table 1). Endonuclease *Eco*RI cleaved pLG214 to produce DNA fragments of 5.9 and 6.4 kilobases, suggesting deletion of 1.5 kilobases from the segment of *ColIdrd-1* present in pLG211 and pLG215. Endonuclease *Bgl*II digests showed that the deletion affects the 4.4-kilobase *Bgl*II fragment found in pLG215 (Fig. 1). This location was confirmed by analysis of

pLG214/pLG215 heteroduplex DNA molecules which demonstrated a deletion loop almost equidistant from the termini generated by cleavage of the plasmids at the unique *Hind*III site (Fig. 2). Heteroduplex DNA formed between endonuclease *Eco*RI-cleaved pLG214 and pLG215 indicated a single internal deletion mapping about 1.4 kilobases from the nearer junction of the cloned DNA and vector (Fig. 2 inset).

Sog⁺ recombinant plasmids specify DNA primase

Table 1 shows that bacteria harbouring *Sog*⁺ recombinant plasmids produce DNA primase. The assay measured the ability

Fig. 4 Glycerol gradient centrifugation of DNA primase specified by pLG214. 47 μ g of protein in 0.2 ml of gradient buffer was layered on a 3.7 ml, 18–43% (wt/vol) glycerol gradient in 20 mM Tris-HCl (pH 7.6), 200 mM NaCl, 1 mM dithiothreitol and 1 mM EDTA. Centrifugation was for 24 h in the Beckman SW-60 rotor at 60,000 r.p.m. Sedimentation was from right to left. 173- μ l fractions were assayed for primase activity (shaded histogram). Aliquots (150 μ l) were subjected to SDS-gel electrophoresis (lower panel). The lane at the right contained 1.9 μ g of primase preparation. Yeast alcohol dehydrogenase (ADH, *M_r* 141,000), *E. coli* alkaline phosphatase (AP, *M_r* 80,000), ovalbumin (OVA, *M_r* 43,000) and lysozyme (LYS, *M_r* 13,970) were run in parallel as markers. Primase was purified from BW86 (pLG214). 280 g of wet cell paste in 50 mM Tris-HCl (pH 7.6), 100 mM NaCl and 5% sucrose was frozen and thawed and then supplemented with EDTA (1 mM), spermidine-3HCl (5 mM), Brij 58 (0.25%) and lysozyme (0.04%) at a final concentration of 1 g cells per 2.5 ml. After 90 min at 0 °C the suspension was centrifuged (90 min, 70,000 g). All subsequent buffers were supplemented with 10% glycerol, 1 mM EDTA and 1 mM dithiothreitol. Supernatant containing 8.3 g of protein was added to a heparin-Sepharose CL6B column and eluted by applying a 600 ml gradient of 0.05–1 M NaCl in 20 mM Tris-HCl (pH 7.6). Primase activity eluted at 0.5 M NaCl. Active fractions (containing 86 mg of protein) were applied to a hydroxyapatite column and protein was eluted by a 300-ml gradient of potassium phosphate pH 7.0 (20–400 mM). Enzyme activity eluted at 150 mM phosphate. Active fractions (42.5 mg of protein) were applied to a phosphocellulose column (Whatman P11) and a 300-ml 0–0.8 M NaCl gradient was applied which eluted primase at 300 mM. After dialysis against buffer A (20 mM Tris-HCl pH 7.6, 50 mM NaCl), 30 ml (8.1 mg of protein) was applied to a DEAE-Sephacel column. Primase activity failed to bind and was washed off with buffer. This fraction (3.9 mg of protein) was applied to a second phosphocellulose column. Active fractions were concentrated by dialysis against 20% polyethylene glycol in buffer A. Yield was 2.6 mg of protein and 80% of activity. The enzyme remains stable at –20 °C for at least 6 months.



of crude cell extracts to stimulate conversion of single-stranded DNA of phage fd to duplex replicative form in a receptor extract containing rifampicin¹⁴. The antibiotic was added to inhibit RNA polymerase, the enzyme naturally required for priming DNA synthesis on viral strands of phage fd¹⁶. A 200-fold higher activity was detected in ColI-containing strains when the plasmid carried the *drd* mutation. This correlates with the requirement for a *drd* mutation in plasmid-mediated suppression of the *dnaG* lesion. *Sog*⁺ recombinant plasmids pLG214 and pLG215 directed synthesis of similar high activities of the primase, indicating that the gene specifying the enzyme is in the same *EcoRI* fragment of *ColI**drd-1* as the gene that causes suppression of *dnaG* mutations.

Polypeptides encoded by *Sog*⁺ plasmids

Plasmids pLG214 and pLG215 segregate into chromosomeless minicells, providing a system for identification of polypeptides encoded by the cloned fragment of *ColI**drd-1* that contains the *sog* gene. Comparison of ³⁵S-methionine-labelled proteins synthesized in pBR325- and pLG215-containing minicells (Fig. 3, lanes a, b) shows that the 7.9-kilobase *EcoRI* restriction fragment of *ColI**drd-1* encodes two polypeptides of *M_r* 240,000 and 180,000. Assuming these to be encoded totally by the *ColI* component of the recombinant plasmid, coding sequences comprising about 0.82 and 0.62 of the DNA, respectively, will be required. The sum of these values implies that the genetic determinants of the two polypeptides share common nucleotides. Lane c indicates that the 6.4-kilobase deletant fragment of *ColI**drd-1* in pLG214 encodes two smaller polypeptides of *M_r* 87,000 and 42,000.

The 240,000 *M_r* protein was also detected in extracts of cells carrying pLG215 or *ColI**drd-1* but not pLG214, pBR325 or *ColI*. Representative results are in Fig. 3, lanes e, g and h. The protein is the largest of the species that stain with Coomassie blue. Lane f contains an extract of BW84 (pLG216). BW84 is the *uvrB*⁺-*chlA*⁺ parent¹³ of BW86 and pLG216 is a *Sog*⁺ plasmid identified after *in vitro* mutagenesis of pLG211 with hydroxylamine¹⁷. The strain was fortuitously isolated during the screening of extracts of appropriate Ap^r BW84 transformants for primase activity. It has the advantage of producing about five times more plasmid primase activity than bacteria harbouring *ColI**drd-1* or pLG215; it also overproduces the 240,000 *M_r* protein.

Purification of a DNA primase and antigenic relationship of polypeptides

A plasmid DNA primase was purified from BW86 (pLG214) using the procedure given in Fig. 4 legend. Chromatography on heparin-Sepharose was very useful because it allowed a 100-fold purification of enzyme in the first step starting with a crude cell extract. The preparation of DNA primase contained one major protein of *M_r* 87,000 which co-sedimented with enzyme activity through a glycerol gradient (Fig. 4). Its sedimentation coefficient was calculated to be 4.2S, indicating that it functions as a single subunit protein. This DNA primase is assumed to be the larger of the two polypeptides detected in pLG214-containing minicells (Fig. 3, lane c).

Using antiserum raised against this DNA primase preparation, immunological studies indicated an antigenic relationship between the two polypeptides encoded by the fragment of *ColI**drd-1* present in pLG215 and the smaller pair of proteins specified by pLG214. The procedure, adapted from that of Towbin *et al.*¹⁸, involved separation of proteins in crude cell extracts by SDS-gel electrophoresis followed by their electrophoretic transfer to a nitrocellulose sheet. The resulting blot was incubated first with antiserum raised¹⁹ in a rabbit against the pLG214-specified primase and then with fluorescein-conjugated goat anti-rabbit immunoglobulins to detect bound antibody.

Figure 5, lanes c and d, show that antiserum reacted with two polypeptides of *M_r* 240,000 and 180,000 in extracts of bacteria harbouring pLG216 or pLG215. In the case of BW84

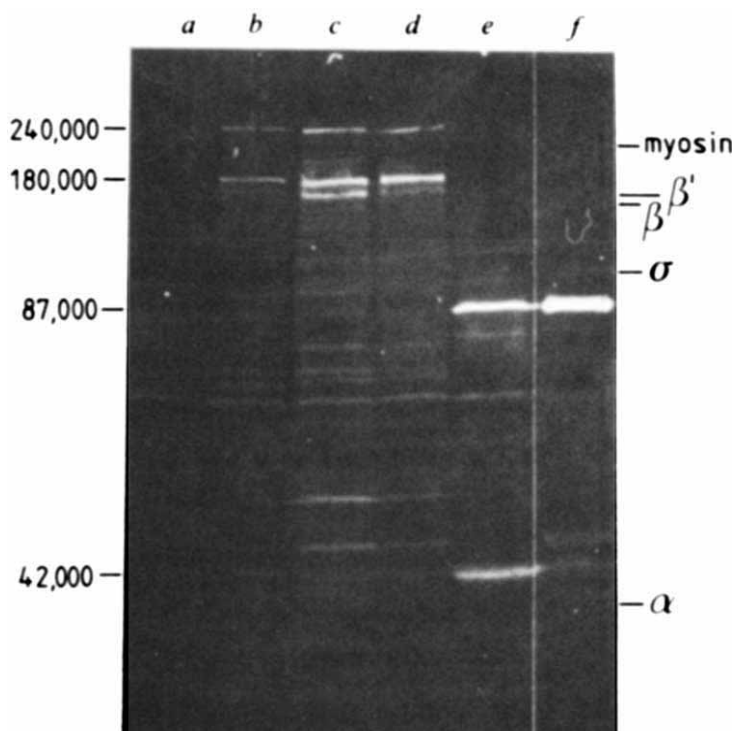


Fig. 5 Immunological analysis of proteins encoded by *Sog*⁺ plasmids. Crude cell extracts were prepared and separated on an SDS-polyacrylamide gel, as described in Fig. 3 legend. Proteins were transferred electrophoretically to a nitrocellulose membrane and reacted with antiserum as described elsewhere¹⁸. Antiserum against purified pLG214-specified DNA primase was applied in a 1:150 dilution. Fluorescein-conjugated goat anti-rabbit IgG were diluted 1:50. Lane a, BW86 (pBR325); b, BW86 (*ColI**drd-1*); c, BW84 (pLG216); d, BW86 (pLG215); e, BW86 (pLG214); f, DNA primase purified from BW86 (pLG214). Migration of molecular weight standards, given under Fig. 3, is shown on the right margin.

(pLG216), the strain that overproduces plasmid primase, a third immunogen (*M_r* 168,000) was also apparent (lane c). The significance of this protein has not been established. Lane e shows that BW86 (pLG214) contains two smaller immunogens (*M_r* 87,000 and 42,000). Only the larger, assumed to be the DNA primase, was present in the antigen (lane f). Preimmune serum was unreactive. Two observations suggest that these polypeptides are encoded solely by the cloned fragment of *ColI**drd-1* and are not products of fused cloned and vector DNA. First, antiserum did not react with protein in an extract of cells harbouring pBR325 (lane a). Second, the 240,000 and 180,000 *M_r* polypeptides are indistinguishable by size from a pair of polypeptides detected in bacteria containing the parental plasmid, *ColI**drd-1* (lane b).

Antiserum, but not preimmune serum, inhibited the ability of crude extracts of BW86 containing either *ColI**drd-1* or pLG215 to promote DNA synthesis on viral strands of phage fd (data not shown). Thus, the primase encoded by these plasmids is structurally related to the 87,000 *M_r* primase of pLG214, the antigen. Similarly, antiserum blocked activity of the primase specified by two other *IncA* plasmids, R144*drd-3* and R64*drd-11*, and the *IncI* γ plasmid, R621a. In contrast, the antibodies failed to inhibit DNA synthesis on viral strands of phages Φ X174 and G4 in an ammonium sulphate fraction¹⁴ of *dna*⁺ cells (data not shown). Thus, the DNA primases encoded by *IncA* plasmids are antigenically unrelated to the *E. coli* enzyme.

Products of the *sog* gene

The following hypothesis is proposed to explain the protein relationships. The *sog* gene encodes a pair of sequence-related

polypeptides of apparent molecular weight 240,000 and 180,000. The deletion mutation in pLG214 disrupts synthesis of the C-terminal segments giving a pair of truncated polypeptides (M_r 87,000 and 42,000). The smaller protein in each pair may be a processed product of the larger or may be generated from a second translational start within the *sog* gene without a change in reading frame. The latter possibility, for which gene *A* of Φ X174 is a precedent²⁰, is favoured by the properties of two *Sog*⁺, primase-defective, amber mutants of pLG211; these specify synthesis of the 180,000 M_r protein but not the 240,000 M_r protein (to be published). It is proposed that DNA primase activity requires part of the N-terminal sequence of the 240,000 and 87,000 M_r proteins. This is consistent with the properties of the amber mutants and the finding that the 87,000 M_r protein, but not the 42,000 M_r protein, purified as a DNA primase. Furthermore, primase activity *in vitro* has been detected in a partially purified preparation of the 240,000 M_r polypeptide (in preparation).

Suppression and DNA primase

The deletion in pLG214 enhances the efficiency of suppression of the *dnaG3* mutation and generates a truncated protein that purifies as a DNA primase. We therefore conclude that plasmid primase is the enzyme responsible for suppression. The enzyme is distinguishable from *dnaG* protein on the basis of molecular weight, antigenic determinants and product of reaction; *in vitro* it directs synthesis of a primer that contains cytidine or cytidine 5'-monophosphate at the 5' terminus (E.L., in preparation) and not an adenosine nucleotide as in the primer synthesized by *dnaG* protein. Thus, the phenomenon of suppression reflects the substitution of an unrelated plasmid enzyme for *E. coli* primase in the discontinuous synthesis of bacterial DNA and any primase-dependent event that may occur at the origin of DNA

replication. The plasmid primase can also substitute *in vivo* during replication of phage λ DNA (unpublished data) and of ColEI-type replicons like the recombinant plasmids used here. Thus DNA primases encoded by IncI α plasmids are remarkable for their ability to use a large variety of templates¹⁴. IncP plasmids, represented by RP4, also specify a DNA primase active in the assay reported here (E.L., unpublished data). This enzyme may also have the potential to act in bacterial DNA replication because derivatives of another IncP plasmid, R68-45, can suppress the *dnaG3* mutation²¹.

IncI α plasmid primase seems to cooperate with DNA polymerase III holoenzyme. Extension of primers *in vitro* requires DNA polymerase III and the *dnaZ* protein¹⁴, which is a component of the holoenzyme²². *In vivo*, bacterial DNA replication in the *dnaZ* mutant, AX727, remains temperature sensitive in the presence of ColI α (unpublished data).

ColI α and R64 α also fail to suppress the temperature sensitivity of DNA replication in *dnaB70* and *dnaB43* mutants^{23,24}, implying that plasmid primase cannot bypass the requirement for *dnaB* protein. The *dnaB* product is thought to act in discontinuous DNA synthesis as a 'mobile promoter' for primer synthesis by bacterial primase⁵. Unlike the latter enzyme, R64 α primase acts independently of *dnaB* protein in the priming reaction on single-stranded DNA of phage Φ X174¹⁴. Thus, the requirement for *dnaB* protein in plasmid primase-mediated DNA replication may reflect involvement of the protein in some process other than the priming reaction—possibly, unwinding of the DNA helix²⁵.

We thank Georg Stöffler for preparing antiserum, Brigitte Geschke and Marianne Schlicht for assistance, Pam Kramer for the electron microscopy, and Alec Jeffreys for his interest. Part of this work was supported by an EMBO short-term fellowship to B.M.W. and SRC grant GR/B24172. We thank H. Schuster for encouraging the collaboration.

Received 4 November 1980; accepted 20 January 1981.

- Ogawa, T. & Okazaki, T. A. *Rev. Biochem.* **49**, 421–457 (1980).
- Lark, K. G. *Nature new Biol.* **240**, 237–240 (1972).
- Rowen, L. & Kornberg, A. *J. biol. Chem.* **253**, 758–764 (1978).
- Wickner, S., Wright, M. & Hurwitz, J. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1613–1618 (1973).
- McMacken, R., Ueda, K. & Kornberg, A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4190–4194 (1977).
- Scherzinger, E., Lanka, E., Morelli, G., Seifert, D. & Yuki, A. *Eur. J. Biochem.* **72**, 543–558 (1977).
- Richardson, C. C. *et al.* *Cold Spring Harb. Symp. quant. Biol.* **43**, 427–440 (1978).
- Nossal, N. *J. biol. Chem.* **255**, 2176–2182 (1980).
- Liu, C. C., Burke, R. L., Hibner, U., Barry, J. & Alberts, B. M. *Cold Spring Harb. Symp. quant. Biol.* **43**, 469–487 (1978).
- Jacob, A. E. *et al.* in *DNA Insertion Elements, Plasmids and Episomes* (eds Bukhari, A. I., Shapiro, J. A. & Adhya, S. L.) 607–638 (Cold Spring Harbor Laboratory, New York, 1977).
- Wilkins, B. M. *J. Bact.* **122**, 899–904 (1975).
- Sasakawa, C. & Yoshikawa, M. *J. Bact.* **133**, 485–491 (1978).
- Boulnois, G. J. & Wilkins, B. M. *Molec. gen. Genet.* **175**, 275–279 (1979).
- Lanka, E., Scherzinger, E., Günther, E. & Schuster, H. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3632–3636 (1979).
- Bolivar, F. *Gene* **4**, 121–136 (1978).
- Brutlag, D., Schekman, R. & Kornberg, A. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2826–2829 (1971).

- Eichenlaub, R. *J. Bact.* **138**, 559–566 (1979).
- Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350–4354 (1979).
- Stöffler, G. & Wittmann, H. G. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2283–2287 (1971).
- Linney, E. & Hayashi, M. *Nature new Biol.* **245**, 6–8 (1973).
- Ludwig, R. A. & Johansen, E. *Plasmid* **3**, 359–361 (1980).
- McHenry, C. & Kornberg, A. *J. biol. Chem.* **252**, 6478–6484 (1977).
- Boulnois, G. J., Beddoes, M. J. & Wilkins, B. M. *J. Bact.* **138**, 324–332 (1979).
- Fenwick, R. G. Jr & Curtiss, R. III *J. Bact.* **116**, 1212–1223 (1973).
- Staudenbauer, W. L., Scherzinger, E. & Lanka, E. *Molec. gen. Genet.* **177**, 113–120 (1979).
- Godson, G. N. & Sinsheimer, R. L. *Biochim. biophys. Acta* **149**, 476–488 (1967).
- Sutcliffe, J. G. *Cold Spring Harb. Symp. quant. Biol.* **43**, 77–90 (1978).
- Jeffreys, A. J. & Flavell, R. A. *Cell* **12**, 429–439 (1977).
- Jeffreys, A. J. *Cell* **18**, 1–10 (1979).
- Daniels, D. L., de Wet, J. R. & Blattner, F. R. *J. Virol.* **30**, 390–400 (1980).
- Sutcliffe, J. G. *Nucleic Acids Res.* **5**, 2721–2728 (1978).
- Timmis, K. N., Cabello, F. & Cohen, S. N. *Molec. gen. Genet.* **162**, 121–137 (1978).
- Davis, R. & Vapnek, D. *J. Bact.* **125**, 1148–1155 (1976).
- Humphries, P. *et al.* *Nucleic Acids Res.* **5**, 905–924 (1978).
- Achtman, M., Manning, P., Edelbluth, C. & Herrlich, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4837–4841 (1979).
- Dougan, G. & Sherratt, D. *Molec. gen. Genet.* **151**, 151–160 (1977).
- Lanka, E., Edelbluth, C., Schlicht, M. & Schuster, H. *J. biol. Chem.* **253**, 5847–5851 (1978).
- Bonner, W. M. & Laskey, R. A. *Eur. J. Biochem.* **46**, 83–88 (1974).

The *E. coli* β -lactamase attenuator mediates growth rate-dependent regulation

Bengtåke Jaurin, Thomas Grundström, Thomas Edlund & Staffan Normark

Department of Microbiology, University of Umeå, S-901 87 Umeå, Sweden

We have identified a new control or attenuator region in the chromosomal β -lactamase operon of *Escherichia coli*. A single base alteration within this attenuator led to a loss in the cell's ability to coordinate its content of β -lactamase with growth rate. We suggest a mechanism through which this mode of regulation operates.

MOST of the more abundant *E. coli* proteins increase in specific amounts with increasing growth rates¹. This group of proteins includes ribosomal proteins, elongation factors, aminoacyl

tRNA synthetases and subunits of RNA polymerase. Also, the amount of rRNA and tRNA increases coordinately with growth rate. Both active and passive regulatory models have been

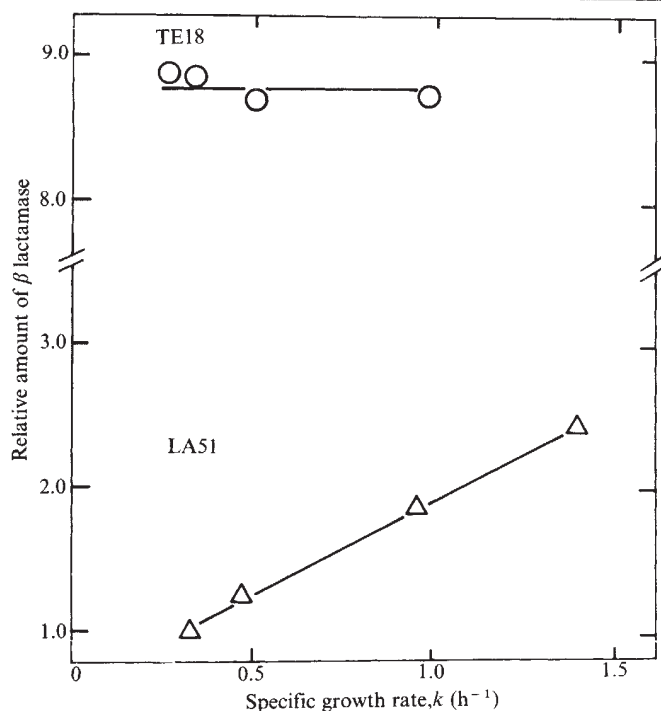


Fig. 1 Relative amount of β -lactamase at different growth rates at 37 °C in *E. coli* K12 strains LA51 (*ampP15G16*) and TE18 (*ampP15G16*, *ampL35A*). The relative amount of β -lactamase per total protein was determined with rocket immunoelectrophoresis as previously described³ and plotted as a function of the first-order constant for growth (k), as calculated from the expression $k = \ln 2/\text{mass doubling time in hours}$. The media used were as follows: medium E + glycerol, medium E + glucose, medium E + glucose + casamino acids, and LB medium³.

proposed to explain the positive correlation between the production of these components of the translation machinery and growth rate².

The relative amount of chromosomally encoded β -lactamase of *E. coli* K12 also increases with growth rate³. This periplasmic enzyme hydrolyses the β -lactam ring of both cephalosporins and penicillins including ampicillin. Its structural gene, *ampC*, has been located at 93.8 min on the *E. coli* linkage map^{4,5}. The strict linear relationship between β -lactamase production and ampicillin resistance enables one to isolate mutants with altered production of β -lactamase by selecting on plates containing different concentrations of ampicillin^{4,6}.

Ampicillin-resistant mutants have been found to carry either *cis*-dominant mutations in a control sequence region denoted *ampA* that is adjacent to *ampC*⁶, or multiple tandem repeats of DNA segments encoding the structural gene⁷. The regulatory region *ampA*, as well as the structural gene *ampC*, have been cloned on a ColE1 hybrid plasmid and by further subcloning localized to a 1,370-base pair DNA segment⁶.

We describe here the sequence and structure of the control region for the β -lactamase gene, *ampC*, in wild-type *E. coli* and two regulatory mutants, and suggest a mechanism whereby the control region regulates transcription in response to the quantity of ribosomes per cell.

A mutation in the control region abolishes the positive correlation between β -lactamase production and growth rate

Strain LA5 is wild type for β -lactamase production. The concentration of ampicillin on broth plates required to kill 50% of the applied wild-type cells (LD_{50}) is $1 \mu\text{g ml}^{-1}$. Strain LA51 is an ampicillin-resistant transductant of strain LA5 (ref. 4). It shows an LD_{50} value of $15 \mu\text{g ml}^{-1}$ of ampicillin, and a corresponding 15-fold increase in β -lactamase production. This strain carries a *cis*-dominant regulatory mutation denoted *ampA1* (ref. 4), which we have recently localized within a

370-base pair DNA segment adjacent to *ampC*⁶. From strain LA51, spontaneous mutants with a further increase in ampicillin resistance have been isolated⁶, one of which, TE18, exhibiting an ampicillin resistance of $62 \mu\text{g ml}^{-1}$ (LD_{50}), carries a second mutation in the same 370-base pair regulatory region for *ampC*.

We have previously demonstrated that the specific amounts of β -lactamase produced both in the wild-type strain LA5 and in the mutant strain LA51 increase proportionally with growth rate³. In contrast, when the growth rate responses of the mutant strains LA51 and TE18 were compared (Fig. 1) it was evident that the growth rate-dependent regulation of β -lactamase had been abolished in the double mutant TE18. Thus, the level of β -lactamase in strain TE18 was much higher than that of LA51,

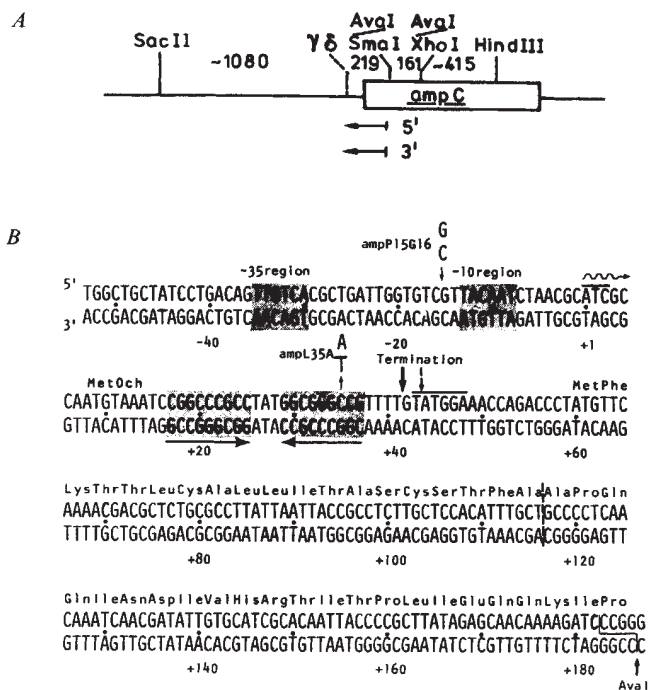


Fig. 2 A, restriction map of the *amp* region and sequencing strategy. The dashed line designates the insertion point of the transposable element $\gamma\delta$ (ref. 6). The box indicates the position of the structural gene, *ampC*. The direction of sequencing from the 5' and 3' ends of DNA cut with *AvaI* (*SmaI*) is indicated by arrows. The 1.5-kilobase *XhoI*-*SacII* fragments from plasmids pNU6, pNU28 and pNU35 were prepared from polyacrylamide gels²¹ and cleaved with *AvaI*. For the 5'-end labelling, 5–30 pmol of 5' ends were incubated in 10 μl of 50 mM Tris-HCl, pH 8.0, with 0.03 units of bacterial alkaline phosphatase at 45 °C for 30 min. 1 μl of 50 mM EGTA was added and the solution was heated at 65 °C for 15 min. 150 μCi vacuum-dried [γ -³²P]ATP (>5,000 Ci mmol⁻¹, Radiochemicals) was dissolved in 9 μl of H₂O and transferred to the DNA mixture. 2.5 μl of 0.5 M Tris-HCl, pH 7.6, 0.1 M MgCl₂, 50 mM dithiothreitol (DTT), 1 mM spermidine, 1 mM EDTA were added and the reaction mixture incubated with 5 units of T4 polynucleotide kinase (Biolabs) at 37 °C for 25 min. After one phenol extraction, labelled DNA was isolated by gel filtration on a Sephadex G-100 column. The 3'-end labelling was in principal as described by Schwartz *et al.*²². However, [α -³²P]dCTP (Radiochemicals) was used, followed by a chase for 10 min at 25 °C with 0.05 mM of cold dCTP. The labelled fragments were cleaved with *Bst*NI (cleaves 625 base pairs to the right of the *SacII* cut). The *AvaI*-*Bst*NI fragments were purified from polyacrylamide gels²³ and subjected to the nucleotide sequencing procedure described by Maxam and Gilbert²³. The degradation products were separated on 0.5-mm thick 8% and 20% sequencing gels with subsequent autoradiography at -20 °C. B, nucleotide sequence of the control region and the N-terminal part of *ampC*. Start of transcription is marked by +1 and the wavy arrow. The major and minor termination points of the short transcript are indicated by vertical solid and dashed arrows, respectively. The stem and loop structure is marked by horizontal arrows and stippled boxes. Solid lines designate possible base pairing with the 3' end of 16S RNA¹⁸. Met and Och designate initiation codon (methionine) and stop codon (ochre), respectively. The limit between the presumed signal peptide and the mature β -lactamase is marked by the vertical dashed line. The positions of the G-C insertion in *ampP15G16* and the A-T transversion in *ampL35A* are represented by vertical arrows. The -35 and -10 promoter regions are marked by stippled boxes.

and it did not change when the steady-state growth rate was varied by a factor of three.

Establishment of the DNA sequences in the wild-type and regulatory mutants

Plasmid pNU6 is a pBR322 derivative which carries the *ampC* gene and the wild-type regulatory region. Plasmids pNU28 and pNU35 carry *ampC* and the mutated control regions of strains LA51 and TE18, respectively⁶. Figure 2A shows a restriction map of a 2.5-kilobase DNA segment containing the β -lactamase gene. We have previously located the regulatory mutations of LA51 and TE18 between the *XhoI* site and the insertion point for the transposable element $-\gamma\delta$ (ref. 6) as shown in Fig. 2A. The nucleotide sequences from the *AvaI* (*SmaI*) site towards the *SacII* site were obtained by both 5'- and 3'-end-labelling and are presented in Fig. 2B. This stretch of DNA was sequenced from plasmids pNU6, pNU28 and pNU35. To localize the beginning of the coding region for *ampC*, the order of the 12 N-terminal amino acids of purified chromosomal β -lactamase was determined by Edman degradation (manuscript in preparation). There was perfect correspondence with the codons from base pairs +117 to +152. The nearest predicted translation start sequence appeared 19 codons before that for the N-terminal amino acid (alanine) of the purified enzyme. These data are consistent with the interpretation that the translational product of *ampC* carries a 19-amino acid N-terminal extension. This stretch of amino acids has all the common structural features of a signal peptide⁸.

Upstream of the transcriptional start base (+1, see below), sequences at nucleotide positions -13 to -8 (-TACAAT-) and -35 to -30 (-TTGTCA-) both show extensive homologies with the conserved -10 and -35 regions of prokaryotic promoters⁹. Also, the sequences between and surrounding these regions show homology with known promoter sequences. When the DNA sequence of the same segment as in pNU6 was determined with the mutant plasmid pNU28, the only difference from the wild type was an insertion of a G·C base pair between positions -16 and -15 (Fig. 2B). Thus, this mutation, denoted *ampP15G16* (previously *ampA1*), which causes a 15-fold increase in β -lactamase synthesis, was located in the presumed *amp* promoter between the conserved -35 and -10 regions.

In the wild-type plasmid pNU6 and in plasmid pNU28, we find a 9-base pair long, exclusively G↔C-containing, dyad symmetry sequence at nucleotide positions +17 to +25 and +29 to +37. This symmetrical DNA sequence is followed by a stretch of four T residues on the noncoding strand. Many terminators for the *E. coli* RNA polymerase have similar sequence features¹⁰. Plasmid pNU35, carrying the *ampC* region of strain TE18, contained in addition to the mutation *ampP15G16*, a C·G to A·T transversion at position +35 (Fig. 2B). This *amp* leader mutation, denoted *ampL35A*, has occurred within the postulated terminator sequence.

In vitro transcription from DNA segments carrying a wild-type or a mutated regulatory region

The 1.5-kilobase *SacII*-*XhoI* DNA fragment which carries the entire sequenced region shown in Fig. 2B, was isolated from plasmids pNU6, pNU28 and pNU35 and added individually to an *in vitro* transcription system. The results are shown in Fig. 3A. From the pNU6 fragment that carries the wild-type regulatory sequence, both a long and a short [α -³²P]CTP-labelled transcript were obtained (L and S in Fig. 3A). The distance from the initiation site for transcription (see below) to the *XhoI* end of the *SacII*-*XhoI* fragment is ~340 base pairs. The long transcript is approximately this length and therefore probably represents a run-off transcript. Furthermore, when the DNA fragment was cleaved at the *AvaI* (*SmaI*) site located 161 base pairs from the *XhoI* site, the 340-base pair long transcript was not found.

Instead, as expected, a new transcript (~180 bases long) appeared (data not shown). The short transcript (S in Fig. 3A) was shown by RNA sequencing analysis to be 41 bases long (see below). The DNA sequence (Fig. 2B and data not shown)

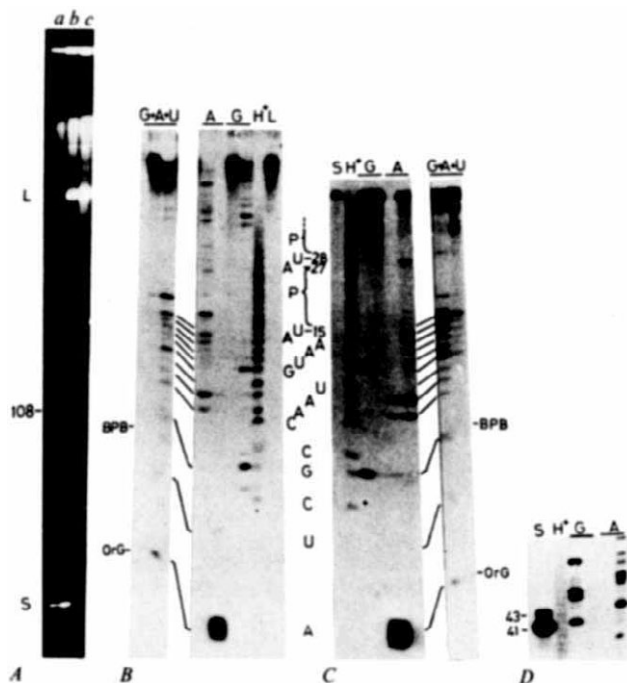


Fig. 3 *In vitro* transcription and RNA sequencing of the *amp* leader region. A shows transcripts obtained from the 1.5-kilobase *XhoI*-*SacII* fragments (see Fig. 2A) carrying different *amp* alleles. The *in vitro* transcription reactions contained 0.25 μ g of the DNA fragment, 2 units of RNA polymerase (Boehringer Mannheim), 100 μ M each of GTP, UTP and ATP, 10 μ M [α -³²P]CTP (1.5 Ci mmol⁻¹) and RNA polymerase binding buffer (5% glycerol, 20 mM Tris-HCl pH 8.0, 10 mM MgCl₂, 0.1 mM EDTA, 0.1 mM DTT and 120 mM KCl) in a total volume of 20 μ l. After 10 min incubation at 37 °C, 8 μ l of formamide-buffer electrophoresis dyes²³ supplemented with 0.4 μ g tRNA were added, and the samples were lyophilized to 8 μ l volume. A 6% acrylamide: bis (30:1) gel with 8 M urea was run²³ and autoradiographed overnight. The position of a 108-base transcript from ColE1²⁴ is indicated. L and S indicate long (340 bases) and short (41 bases) transcripts, respectively. The sources of the DNA fragments used were pNU6 (*ampP*⁺, *ampL*⁺) (lane a), pNU28 (*ampP15G16*, *ampL*⁺) (lane b) and pNU35 (*ampP15G16*, *ampL35A*) (lane c). B and C show autoradiographs of RNA sequencing gels of the long and short transcripts, respectively. The 5'-end-labelled transcripts were synthesized in RNA polymerase binding buffer supplemented with 150 μ M CTP and UTP, 40 μ M GTP, 40 μ M [γ -³²P]ATP (100 Ci mmol⁻¹) and 10 units RNA polymerase in a total volume of 100 μ l. The long transcript was synthesized with 2.2 μ g of the 1.3-kilobase *SmaI*-*SacII* fragment from pNU35 as template, and the short transcript with 1.25 μ g of the 1.5-kilobase *XhoI*-*SacII* fragment from pNU28. After incubation at 37 °C for 15 min, the samples were treated as above except that the gel contained 8% acrylamide. The transcripts were isolated by the method of Maxam and Gilbert²³ and the RNA sequencing reactions were as described by Krupp and Gross²⁵. All reactions were done with 4 μ g of carrier tRNA. The G + A + U reactions were carried out with 5 and 50 μ g of *Neurospora* endonuclease, respectively, the A reactions with 0.05 and 0.5 units of RNase U₂, respectively, and the G reactions with 1 and 10 units of RNase T₁, respectively. H' indicates controlled partial acid hydrolysis giving labelled mRNA of all sizes. L and S indicate untreated long and short transcripts, respectively. The position of the orange G and bromophenol blue markers which migrate with 2 and 6 bases length, respectively²⁶, on the 25% polyacrylamide-urea gels used is marked OrG and BPB. Note that due to the absence of a phosphate at their 3' ends, oligonucleotides produced by *Neurospora* nuclease move more slowly in the gel²⁵. These bands have been connected with the corresponding bands from the other reactions. The sequence of the transcripts is indicated between B and C. P indicates positions where the transcripts due to their secondary structure are protected from the action of the enzymes. D is an autoradiograph of a part of an 8% polyacrylamide gel, showing the sequence of the long transcript around the termination point of the short transcript. Note that in contrast to the transcripts, the labelled oligonucleotide degradation products have a phosphate at their 3' end. Consequently, when interpreting D it must be remembered that the untreated ladder is displaced upwards by about half a nucleotide unit. Most of the short transcripts terminate at base 41 of the long transcript with minor termination at base 43.

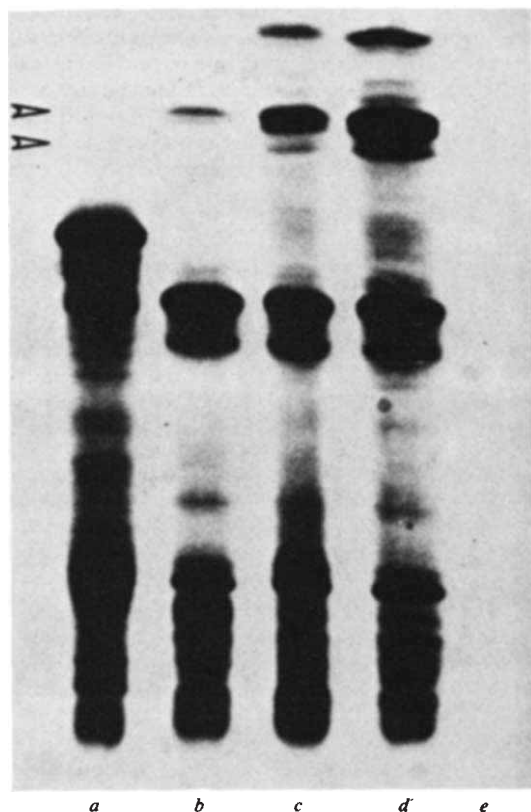


Fig. 4 Expression of β -lactamase in a coupled *in vitro* transcription-translation system. The system is originally derived from Zubay²⁷, and modified by Johnsen²⁸ and Collins²⁹. The polypeptides were separated on a 10–17.5% gradient polyacrylamide gel as previously described⁶. The gel was fluorographed using En³hance (NEN). The plasmids used were: a, pBR322; b, pNU6; c, pNU28; d, pNU35; e, no plasmid DNA added. The 36K and 38K proteins which represent mature β -lactamase and presumed precursor β -lactamase, respectively, are marked by arrows.

reveals that the long transcript contains seven times as many C residues as the short one. Taking this into account, and with the relative band intensities obtained by scanning of the autoradiographs, we conclude that the short transcript is reproducibly about 15 times more abundant than the longer one.

In vitro transcription with the same amount of the analogous DNA fragment from pNU28 (carrying the *ampP15G16* mutation) resulted in an increase in the synthesis of both the long and the short transcripts (Fig. 3A, lane b). This contrasts with the result obtained when the corresponding *SacII*–*XhoI* fragment from pNU35 (carrying both the *ampP15G16* and *ampL35A* mutations) (lane c) was transcribed; no synthesis of the short 41-base transcript was observed. Moreover, the amount of the long run-off transcript was increased. The loss of the short transcript associated with an increase in the amount of the long run-off transcript suggests that the second mutation in TE18, *ampL35A*, abolishes termination in the *amp* leader region.

Localization of initiation and termination points for transcription

The initiation nucleotide *in vitro* of β -lactamase mRNA was determined by transcription using 10 μ M of one nucleotide triphosphate (NTP) and 150 μ M of the other three NTPs. Synthesis of β -lactamase transcripts was severely inhibited by ATP limitation but not by limitation of any of the other three NTPs (data not shown). This suggests that A is the 5' nucleotide. The long run-off transcript and the short transcript were labelled

using [γ -³²P]ATP, and subjected to RNA sequencing by partial enzymatic degradation (Fig. 3B, C). The first 16 nucleotides in the two transcripts were identical and homologous to the DNA sequence from bases +1 to +16. This shows that the A at +1 is the start base for both these mRNAs and confirms the position for the β -lactamase promoter. The nucleotides in the RNA sequence between positions 17 and 37 except for the A and U residues at positions 27 and 28, respectively, were protected against enzymatic cleavage. As the enzymes used are specific for single-stranded RNA¹¹, this finding implicates the existence of a 9-base pair RNA stem with a 3-base loop at positions 26–28. The point of termination of the short transcript was determined by comparing the length of non-degraded short transcript with that of enzymatically degraded long transcript (Fig. 3D). At least 99% of the short transcript ended with the U at position 41. A minor fraction terminated two nucleotides later.

Expression of the *amp* operon in a coupled *in vitro* transcription-translation system

When DNA from plasmids pNU6, pNU28 and pNU35 were added to a coupled transcription-translation system a protein with an apparent molecular weight (MW) of 38,000 (38K) was produced. The relative production of this protein from the respective plasmids was ~1:15:60 (Fig. 4). Purified β -lactamase migrates with an apparent MW of 36K. A faint band appeared at this position in the gel when plasmid pNU6 was used. In the mutants the corresponding protein band was more intense. We suggest that the 38K protein observed is the precursor form of the *E. coli* chromosomal β -lactamase postulated on the basis of the DNA sequence. The uppermost band with an apparent MW of 56K also increased in intensity in the mutants—the nature of this band is not known.

The fourfold difference between plasmid pNU28 and pNU35 in production of the pre- β -lactamase in the coupled transcription-translation system, corresponds to the fourfold differences in steady-state levels of β -lactamase found *in vivo*. On the other hand, in the *in vitro* transcription system the difference in transcriptional readthrough was about 16-fold. This suggests that a factor(s) present *in vivo* and in the coupled

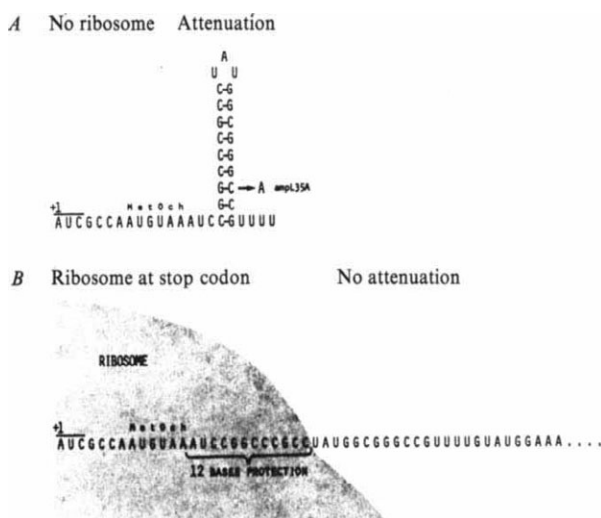


Fig. 5 A model for growth rate-dependent attenuation of transcription. The stippled area represents a ribosome. The symbols are the same as in Fig. 2B. A, terminated transcript; B, hypothetical protection by the ribosome preventing formation of the termination stem, leading to readthrough (see text for discussion).

system increases the transcriptional readthrough about fourfold compared with a pure *in vitro* transcription system.

Features of the β -lactamase attenuator

One of our major conclusions is that the *amp* operon contains within its leader region a site for termination of transcription that is recognized by the *E. coli* RNA polymerase *in vitro*. Factors such as the rho protein¹² are not required for efficient termination *in vitro* at this site. However, we do not know whether the rho protein has any effect in the *in vivo* situation. On the other hand, the terminator in the *amp* leader has a structure similar to those of sequenced rho-independent terminators¹⁰.

The base pair substitution *ampL35A* occurred 7 base pairs upstream from the major termination site and within the 9-base pair symmetrical DNA sequence (Fig. 2B). From both thermodynamic considerations and the relative susceptibility to enzymatic cleavage, we suggest that the leader RNA transcript forms a secondary structure with a 9-base pair G $\leftrightarrow$ C-containing stem and a 3-base loop (Fig. 5A). The *ampL35A* mutation decreases the thermodynamic stability of such an RNA stem from a ΔG value of -29.4 kcal to -17.4 kcal (ref. 13). Termination of transcription *in vitro* decreases to less than 1% in template with the *ampL35A* mutation. It seems likely that the terminator in the *amp* leader also operates *in vivo*, because β -lactamase synthesis at the maximal growth rate is roughly four times higher in the *ampL35A*-carrying mutant strain, TE18, than in its parent strain, LA51.

Growth rate-dependent regulation of *ampC* is exerted at the level of anti-termination

In the *trp*, *phe*, *his*, *leu*, *ilv* and *thr* biosynthetic operons, a potential coding region for a short leader peptide precedes a terminator¹⁴⁻¹⁶. In each case, the leader peptide contains 'regulatory' codons for the amino acids that regulate the operon. In the model originally put forward by Lee and Yanofsky¹⁷ it is proposed that synthesis of the leader peptide is involved in the regulation of transcription termination at the attenuator. This regulation depends on the formation of mutually exclusive RNA secondary structures. Ribosomes translating the entire coding region would permit formation of the terminator secondary stem and loop structure, whereas ribosomes slowed down at the 'regulatory' codons, due to a deficiency in the concentration of the charged cognate tRNA, would favour formation of an alternative structure and thereby preclude formation of the terminator structure.

The 41-base long *amp* leader transcript is considerably shorter than other known leader transcripts and has no potential coding region for a leader peptide. Moreover, the *amp* leader RNA cannot form any alternative secondary structure that precludes formation of the terminator stem and loop. However, examination of the RNA sequence reveals that the first three bases are complementary to a sequence near the 3' end of 16S rRNA (ref. 18). Furthermore, bases 8-13 consist of an initiation codon (AUG) that is followed directly by an ochre stop codon (UAA). Thus, the *amp* leader RNA has a potential ribosome binding site and may form a translation initiation complex. Our major concerns are whether or not this potential ribosome binding site is used and whether it could have a regulatory effect on the termination of transcription. A ribosome stalled at the stop codon would protect ~ 12 bases downstream¹⁹. As shown in Fig. 5B this could prevent formation of the termination stem and loop structure.

Expression from the *amp* operon is positively correlated with the steady-state growth rate. In theory, this growth rate-dependent regulation could occur at the level of initiation or

termination of transcription or both. The *ampP15G16* mutation increases transcription initiation *in vitro* by a factor of 15 without significantly affecting the degree of transcription termination. As this up-promotor mutation did not affect the growth rate-dependent expression of *ampC* we find it unlikely that the *amp* promoter participates in this regulation.

In the double mutant, TE18, expression of *ampC* has become independent of the growth rate. This fact, together with the observation that the mutation *ampL35A* abolishes termination of transcription *in vitro* at the *amp* attenuator, suggests that the growth rate-dependent regulation of the *amp* operon is exerted at the level of anti-termination. The anti-termination factor in our system is not known. We hypothesize that ribosome binding at the *amp* leader RNA has a regulatory role. The number of ribosomes per cell mass increases, over a broad range, proportionally with growth rate². If binding of ribosomes to the *amp* leader region is proportional to the cellular concentration of ribosomes, the degree of anti-termination would increase proportionally with growth rate.

Many proteins exhibit the same growth rate-dependent regulation as the chromosomal β -lactamase. For many operons this mode of regulation might also be exerted at the level of anti-termination. Terminator-like sequences have been found in the leader DNA of the *rpsE* and *rpsL* operons. From the *rpsL* operon, a short transcript has been obtained²⁰ which suggests that termination can at least occur *in vitro*. It is, however, not known if these possible terminator sequences are involved in the regulation of the operons.

The regulatory region of *ampC* consists of two control regions, a promoter and an attenuator. The up-promotor mutation described here changes the distance between the two conserved regions in the promoter. Possibly this is of importance for efficient RNA polymerase recognition. Because of the ease with which one can isolate and analyse mutations in the regulatory regions of the *amp* operon, this system can be used to increase our understanding of the functions of promoters and attenuators.

We thank Torbjörn Nilsson and Per Pettersson for help with the amino acid sequence determination, Monica Persson for technical assistance, and Stewart Cole, John K. Davies and Christopher Korch for careful reading of the manuscript. This work was supported by grants from the Swedish Natural Science Research Council (Dnr 3373) and from the Swedish MRC (Dnr 5428).

Received 28 July; accepted 24 December 1980.

- Pedersen, S., Block, P. L., Reeh, S. & Neidhardt, F. C. *Cell* **14**, 179-190 (1978).
- Maaloe, O. in *Biological Regulation and Development* Vol. 1 (ed. Goldberger, R. F.) 487-542 (Plenum, New York, 1979).
- Jaurin, B. & Normark, S. *J. Bact.* **138**, 896-902 (1979).
- Normark, S. & Burman, L. G. *J. Bact.* **132**, 1-7 (1977).
- Bachmann, B. J. & Low, B. K. *Microbiol. Rev.* **44**, 1-56 (1980).
- Grundström, T., Jaurin, B., Edlund, T. & Normark, S. *J. Bact.* **143**, 1127-1134 (1980).
- Edlund, T., Grundström, T. & Normark, S. *Molec. gen. Genet.* **173**, 115-125 (1979).
- Inoué, M. & Halegoua, S. *CRC crit. Rev. Biochem.* **7**, 339-371 (1980).
- Siebenlist, U., Simpson, R. B. & Gilbert, W. *Cell* **20**, 269-281 (1980).
- Rosenberg, M. & Court, D. A. *Rev. Genet.* **13**, 319-353 (1979).
- Peattie, D. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1760-1764 (1979).
- Pribnow, D. in *Biological Regulation and Development* Vol. 1 (ed. Goldberger, R. F.) 219-277 (Plenum, New York, 1979).
- Tinoco, I. Jr et al. *Nature new Biol.* **246**, 40-41 (1973).
- Keller, E. & Calvo, J. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6186-6190 (1979).
- Nargang, F. E., Subrahmanyam, C. S. & Umbarger, H. E. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1823-1827 (1980).
- Lawther, R. P. & Hatfield, G. W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1862-1866 (1980).
- Lee, F. & Yanofsky, C. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4365-4369 (1977).
- Shine, J. & Dalgarno, L. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1342-1346 (1974).
- Steitz, J. in *Biological Regulation and Development* Vol. 1 (ed. Goldberger, R. F.) 349-399 (Plenum, New York, 1979).
- Post, L. E., Arfsten, A. E., Reusser, F. & Nomura, M. *Cell* **15**, 215-229 (1978).
- Akujärvi, G. & Pettersson, U. *Virology* **91**, 477-480 (1978).
- Schwartz, E., Scherer, G., Hobom, G. & Kössel, H. *Nature* **272**, 410-414 (1978).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
- Morita, M. & Oka, A. *Eur. J. Biochem.* **97**, 435-443 (1979).
- Krupp, G. & Gross, H. J. *Nucleic Acids Res.* **6**, 3481-3490 (1979).
- Horii, T., Ogawa, T. & Ogawa, H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 313-317 (1980).
- Zubay, G. A. *Rev. Genet.* **7**, 267-287 (1973).
- Johnsen, M. thesis, Univ. Copenhagen (1976).
- Collins, J. *Gene* **6**, 29-42 (1979).

LETTERS

IUE detection of bursts of H Ly α emission from SaturnJ. T. Clarke*, H. W. Moos* $\S$, S. K. Atreya $\dagger$ & A. L. Lane $\ddagger$

* The Johns Hopkins University, Baltimore, Maryland 21218, USA

 $\dagger$ University of Michigan, Ann Arbor, Michigan 48109, USA $\ddagger$ Jet Propulsion Laboratory, Pasadena, California 91103, USA

An extended source of H Ly α emission (1,216 Å) from the vicinity of Saturn has been detected previously by a rocket-borne spectrometer¹, the OAO-Copernicus satellite², and the Pioneer 11 spacecraft³. During the rocket experiment, in which Saturn was observed on 15 March 1975, 700 R $\pm$ 50% of Ly α emission was detected from the planet and an additional 200 R $\pm$ 50% from a 53 arc s diameter circular area centred on the planet. This additional emission was believed to emanate from the ring system. The Copernicus observations, carried out in April 1976 and April 1977, yielded Ly α intensities of 1.40 $\pm$ 0.45 kR for the saturnian disk, <100 R for the rings, and 200 $\pm$ 100 R for Titan's torus. During the Pioneer 11 Saturn flyby in August–September 1979, the long-wavelength channel UV photometer detected emissions from the planet, the vicinity of the rings, and the orbital path of Titan; these emissions were interpreted as resonant scattering of solar Ly α radiation by H atoms and possibly auroral Ly α emission from the planet. Several mechanisms for producing H atoms outside of Saturn's atmosphere have been proposed, including sputtering of water ice from the rings by magnetospheric charged particles⁴, 'photo-sputtering' of the ring water ice by solar UV radiation⁵, and escape of H atoms from Titan's atmosphere⁶. A more recent suggestion is the possible impact on the rings and subsequent neutralization of the protons from an extended saturnian ionosphere¹⁵. We report here a new investigation of these potential sources of Ly α emission in a series of observations of the saturnian system carried out between January and July 1980 using the short wavelength (SWP) spectrograph of the IUE Observatory^{7,8}. North-south maps of the Ly α emission across the planet disk have shown pronounced spatial asymmetries in emission brightness. These asymmetries vary markedly on a time scale of days and are interpreted as bursts of Ly α emission of as much as 1 kR brightness averaged over a 6 $\times$ 10 arc s area, above a constant planetary emission level of 700–800 R. In fact, the Ly α emission peaks appear as essentially point source features in these data: if the emitting region is smaller than the 6 $\times$ 10 arc s instrumental resolution, the surface brightness must be proportionally higher.

Spatial imaging with the IUE has been accomplished both in a series of exposures of different regions of the saturnian system and by sampling within the field-of-view of the large spectrograph entrance aperture in single exposures. The image of the field-of-view subtended by the aperture is focused directly onto the detector at each wavelength; because the angular resolution of the instrument is 5–6 arc s on planetary objects and the large aperture extends 10 $\times$ 23 arc s, imaging perpendicular to dispersion is possible in a single exposure. By moving the $\sim$ 16 arc s diameter image of Saturn 5–10 arc s between exposures along the major axis of the aperture (which is constrained to point roughly north-south), maps of the disk emission plus the background geocoronal and interplanetary Ly α emission immediately north and south of Saturn have been obtained with 6 arc s spatial resolution. In each of the exposures centred on the polar regions, therefore, the level of background Ly α is directly measured and subtracted from the planetary emission; in exposures centred on Saturn the background level has been

determined by matching the level of planetary emission to that observed in the polar exposures immediately preceding or following. A relative sensitivity correction in the direction perpendicular to dispersion has been derived from exposures of the background Ly α emission and applied to the data reported here. The error in applying this correction has been determined by testing it on other exposures of background Ly α emission and measuring the average deviation from the mean level of emission; this error is $\pm$ 200 R. The pointing of the spacecraft during the north-south maps has been determined by noting the position in the aperture of the sharp drop in grating-scattered long wavelength radiation at the northern and southern edges of the saturnian disk. The north-south positioning of the aperture is believed to be accurate to within 1–2 arc s. This same technique of north-south mapping has been used in observations of the polar auroral Ly α emission from Jupiter⁹.

Representative north-south maps of the Ly α emission are plotted in Fig. 1 from observations performed on 19 January, 5 May and 22 July (all 1980). Non-equatorial peaks in the emission are apparent in the data from the first two days, while the July data are symmetric north-south with an equatorial brightness of 800 R. This equatorial brightness in the individual exposures on 22 July was measured over one-half of a rotation of the planet; therefore, 700–800 R is interpreted as a constant level of planetary emission from Saturn's upper atmosphere. The data of 22 July provide a good fit to a model emission profile

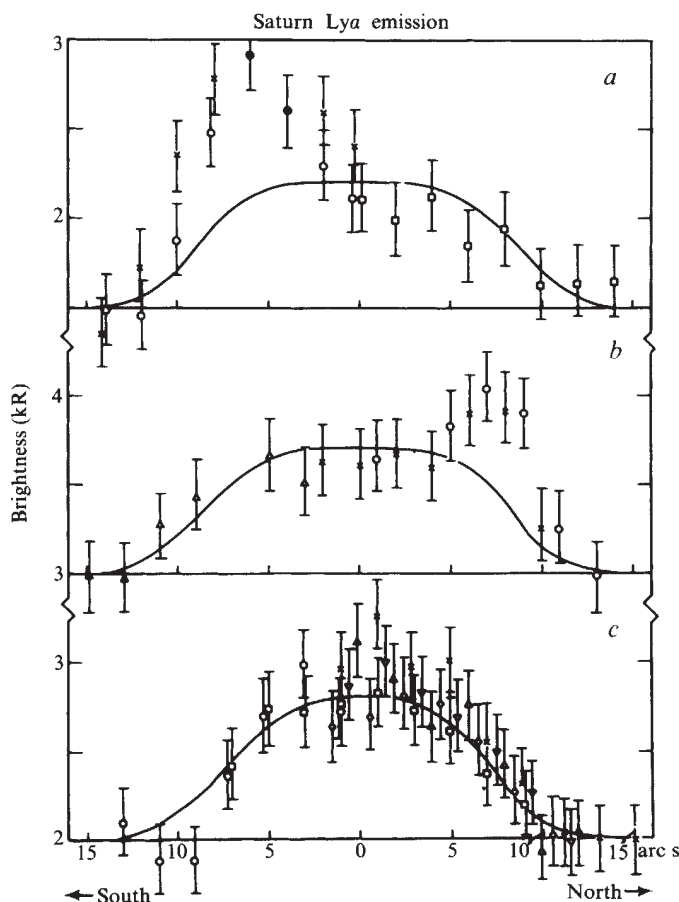


Fig. 1 North-south maps of the H Ly α emission from Saturn plus sky background. Different point characters represent different camera exposures, and the $\pm$ 200 R error bars are discussed in the text. The measured background varied by as much as $\pm$ 500 R in a series of exposures, and the exposures plotted have been normalized to the same level of background (set at the average background level observed on a given day). *a*, 19 January 1980; *b*, 5 May 1980; *c*, 22 July 1980.

$\S$ Permanent address: Joint Institute for Laboratory Astrophysics and Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado, USA.

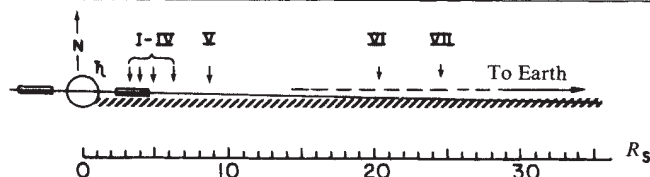


Fig. 2 Observing geometry during the exposures of 19 January 1980. The dashed line and arrow indicate the line-of-sight from the Earth, and the solid line represents the ring plane (which is tilted 1.6° south of the line-of-sight). The enhanced $\text{Ly}\alpha$ emission observed apparently from the south pole of Saturn could actually have come from anywhere in the shaded region. Roman numerals indicate the orbital radii of Saturn's inner moons (VI is Titan).

(the solid line in Fig. 1) representing a uniformly emitting disk of 800 R brightness smoothed by the 6 arc s instrumental resolution. Note that the decrease in emission from the equator to the poles in the model profile is due to the spatial resolution of the instrument, not limb darkening. If this emission is completely produced by resonant scattering of solar $\text{Ly}\alpha$ from upper atmospheric H it corresponds to an H column abundance on the order of 10^{14} cm^{-2} , following the simple resonant scattering model for Jupiter outlined by Clarke *et al.*¹⁰. This model assumes an exospheric temperature of 1,200 K, a solar flux at line centre at the Earth of $5.14 \times 10^{11} \text{ photons cm}^{-2} \text{ s}^{-1} \text{ \AA}^{-1}$ (ref. 11), and no absorption by interplanetary H along the path to and from Saturn. The assumed temperature value is uncertain, but 1,200 K is consistent with the Pioneer 11 radio occultation data¹². A disk-averaged $\text{Ly}\alpha$ intensity of 700–800 R implies that a homopause value of the eddy diffusion coefficient is in excess of $10^5 \text{ cm}^2 \text{ s}^{-1}$. Uncertainties about the role of particle precipitation and the methane density distribution and temperature structure of the mesosphere and thermosphere prevent an exact calculation of the eddy diffusion coefficient at this time. Interplanetary absorption of $\text{Ly}\alpha$ on the round trip to Saturn has been modelled by Bertaux *et al.*¹³, who predict 40% absorption at the time of these observations; thus, Saturn's $\text{Ly}\alpha$ emission could be significantly brighter as measured by a spacecraft in the vicinity of the planet. An H column of 10^{14} cm^{-2} is optically thick and Saturn should exhibit limb darkening at $\text{Ly}\alpha$; however, with the existing spatial resolution it is not possible to distinguish between a cosine darkening function and a uniformly emitting disk. Note that the points plotted in Fig. 1 represent average brightness over an area extending 10 arc s in the east-west direction; assuming a cosine darkening east-west (as well as north-south), the actual sub-solar point brightness would be $\sim 5\%$ higher than that measured.

Exposures were also taken of Titan near western elongation on 5 May 1980, of the sky background 4 arc min north of Saturn on 9 May 1980, and of the region of the outer rings 20 arc s east of the centre of Saturn on 9 May 1980. To within ± 200 R no $\text{Ly}\alpha$ emission above the level of sky background was detected from either Titan or the rings, although the ring exposure was taken on a day when no bursts of $\text{Ly}\alpha$ emission above the 700–800 R level were detected on the saturnian disk. The additional sky background exposure was taken to check for a possible extended H atmosphere around the saturnian system, as hypothesized by McDonough and Brice⁶; however, the background brightness 4 arc min north of Saturn was the same within ± 200 R as that measured immediately north and south of Saturn in the preceding and following exposures. All exposures to date are listed in Table 1, specifying the target area, the longitude of the central meridian of Saturn, the tilt angle of the ring plane north or south of the line of sight from Earth, and the observed level of emission above the 700–800 R planetary level. The longitudes listed are in the Saturn longitude system (λ_{SLS}) defined by the recent observations by the Voyager Planetary Radio Astronomy Group¹⁴. A comparison was also made between the data in Table 1 and the occurrence of radio emission from Saturn, but no direct correlation was discovered.

Unfortunately, the angular resolution of the IUE is not sufficient to pin down the source of the enhanced $\text{Ly}\alpha$ emission.

The excess emission from the southern half of the saturnian disk observed on 19 January, for example, could have originated anywhere along the shaded region shown in Fig. 2 and satisfied our viewing constraints. However, because the rings were within $\pm 2^\circ$ of being edge-on to the Sun throughout these observations and the $\text{Ly}\alpha$ bursts were variable on a time scale of days, it seems unlikely that photo-sputtering of ring particle ice is a plausible source mechanism for these bursts. In addition, <200 R of $\text{Ly}\alpha$ emission was seen from the vicinity of Titan on the same day that a burst of brightness 600 R was detected from the northern half of the saturnian disk, implying that the source of this enhanced emission was not Titan or an associated torus. The most plausible production mechanisms seem to be either: (1) aurora in Saturn's upper atmosphere or (2) a cloud of H atoms in the vicinity of the rings which are released by charged particle sputtering and radiate through resonant scattering of solar $\text{Ly}\alpha$ radiation (and possibly through charged particle excitation). More accurate knowledge of the source of these bursts of $\text{Ly}\alpha$ emission requires better spatial resolution and more flexibility in the angle of view, as will be available from the Voyager 1 and 2 spacecraft now approaching encounter with Saturn.

Table 1 Observations of H $\text{Ly}\alpha$ emission from Saturn

Date (1980)	Target	Ring plane tilt angle	λ_{SLS} (CML)	Average background level (kR)	Emission above model level (R)
19 January	5° S	1.6° S	319°	1.5	+900
	5° S	1.6° S	346°	1.5	+900
	5° N	1.6° S	34°	1.5	—
12 March	8° N	0°	306°	2.5	—
	Centre	0°	338°	2.5	—
13 March	8° S	0°	10°	2.5	—
3 May	8° N	1.5° N	315°	1.4	+300
	8° S	1.5° N	350°	1.4	+300
	Centre	1.5° N	25°	1.4	+300
5 May	8° N	1.5° N	355°	3.0	+600
	Centre	1.5° N	32°	3.0	—
	8° S	1.5° N	75°	3.0	—
	Titan	1.5° N	117°	3.0	—
9 May	8° N	1.5° N	76°	1.8	—
	4° N	1.5° N	112°	1.8	—
	8° S	1.5° N	149°	1.8	—
	20° E	1.5° N	189°	1.8	—
22 July	7° N	0°	118°	2.0	—
	7° S	0°	155°	2.0	—
	7° N	0°	189°	2.0	—
	Centre	0°	223°	2.0	—
	7° N	0°	262°	2.0	—
	7° N	0°	296°	2.0	—

We thank the IUE Observatory staff for acquisition and reduction of the satellite data. This research was supported by NASA under grant NSF 5393 to the Johns Hopkins University. S.K.A. acknowledges support received from a NASA-Planetary Atmospheres Grant to the University of Michigan.

Received 16 October; accepted 4 December 1980.

- Weiser, H., Vitz, R. C. & Moos, H. W. *Science* **197**, 755–757 (1977).
- Barker, E. S., Cazes, S., Emerich, C., Vidal-Madjar, A. & Owen, T. *Astrophys. J.* **242**, 383–394 (1980).
- Judge, D. L., Wu, F.-M. & Carlson, R. W. *Science* **207**, 431–434 (1980).
- Cheng, A. F. & Lanzerotti, L. J. *J. geophys. Res.* **83**, 2597–2602 (1978).
- Carlson, R. W. *Nature* **283**, 461 (1980).
- McDonough, T. R. & Brice, N. M. *Icarus* **20**, 136–145 (1973).
- Boggeess, A. *et al. Nature* **275**, 377–385 (1978).
- Lane, A. L. *et al. Nature* **275**, 414–415 (1978).
- Clarke, J. T., Moos, H. W., Atreya, S. K. & Lane, A. L. *Astrophys. J. Lett.* **241**, 179–182 (1980).
- Clarke, J. T. *et al. Astrophys. J.* **249**, 696–701 (1980).
- Mount, G. H., Rottman, G. J. & Timothy, J. G. *J. geophys. Res.* **85**, 4271–4274 (1980).
- Kliore, A. J. *et al. Science* **207**, 446–449 (1980).
- Bertaux, J. L., Festou, M., Barker, E. S. & Jenkins, E. B. *Astrophys. J.* **238**, 1152–1159 (1980).
- Desch, M. D. & Kaiser, M. L. (in preparation).
- Luhmann, J. G. & Walker, R. J. *J. geophys. Res.* (in the press).

Mechanical aggregation of enstatite chondrites from an inhomogeneous debris cloud

C. A. Leitch & J. V. Smith

Department of the Geophysical Sciences, University of Chicago, Chicago, Illinois 60637, USA

Enstatite chondrites have oxygen isotope ratios closer to those of the Earth and Moon than other meteorites. Their minerals are chemically reduced; metal contains Si, and some Ca, Ti, Mg and Mn are incorporated in sulphides rather than silicates. Clinostite and olivine are virtually Fe-free^{1,2}. We have reported³ two types of clinostite in the Indarch enstatite chondrite, one luminescing blue and one red. Similar clinostites in the Kota-Kota enstatite chondrite are associated with two distinct types of forsteritic olivine, one luminescing orange and the other blue. The textural relations and differences in chemical composition cannot be explained by progressive condensation from the solar nebula and require the mechanical mixing of material from at least two sources. We suggest here that enstatite chondrites result from mechanical and chemical processes during aerodynamic sorting and gravitational settling of debris from a hot cloud of dust, liquid and gas produced during collision of planetesimals.

Kota-Kota, like Indarch, contains four types of clinostite-rich chondrules: porphyritic > fine-grained > barred > radiating. One radiating chondrule composed primarily of SiO₂ was observed. A luminescence microscope revealed distinct red and blue clinostite grains, as well as orange and blue olivine grains. The latter were not found in the first thin section of the Indarch chondrite³, but a few orange and blue grains were found in six new sections. Porphyritic chondrules are extremely

complex and a single chondrule may contain all four phases (Fig. 1). Red clinostite grains may enclose blue clinostite, orange olivine and/or blue olivine, while individual blue clinostite grains may enclose orange and/or blue olivine. Red clinostite was not observed inside blue clinostite. Fine-grained chondrules contain blue clinostite and rare blue olivine. Both barred and radiating chondrules are composed predominantly of lamellae of either red or blue clinostite. All chondrule types are surrounded by a shell composed of a fine-grained mixture of all minerals found in the meteorite. Similar shells were found surrounding chondrules in Indarch³ and in ordinary chondrites⁴.

Electron microprobe analyses were obtained with an ARL-EMX instrument. Magnesium, Si and Fe were determined with a solid state detector at 15 kV, 0.1 μ A and 120 s live time. Minor elements were determined with crystal spectrometers and 25 kV and 2 μ A. Counting times were selected for peaks and backgrounds to achieve a precision near 30 p.p.m. (1 σ) for individual analyses. Standards were: Ca₂P₂O₇(P), Di₈₅Jd₁₅(Ca, Al, Na), Corning glass V(Ti, Cr), Corning glass W(Mn) and Corning glass X(Ni).

As for Indarch³, minor element contents correlate with luminescence colour and not with chondrule type. Representative analyses in Table 1 are taken only from the chondrule in Fig. 1. Table 2 shows observed ranges in minor element contents for both olivine and clinostite. Red clinostites have

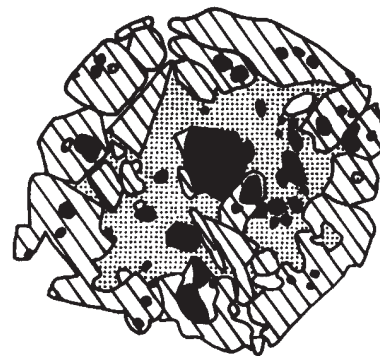


Fig. 1 Schematic representation of a porphyritic chondrule. Black, clear, hatched and stippled areas are respectively, orange olivine, blue olivine, red clinostite and blue clinostite. Blue clinostite is polycrystalline, but individual grain boundaries are not shown. The chondrule is 410- μ m wide.

Table 1 Olivine and pyroxene analyses from one Kota-Kota chondrule

	1	2	3	4	5	6
P ₂ O ₅	0.005	0.004	0.005	0.005	NA	<0.1
SiO ₂	42.3	42.2	58.9	58.9	58.9	58.2
TiO ₂	0.016	0.024	0.079	0.008	NA	0.03
Al ₂ O ₃	0.051	0.077	0.39	0.108	0.27	0.32
Cr ₂ O ₃	0.23	0.064	0.32	0.024	NA	0.36
FeO _T	0.96	0.86	0.95	0.81	0.87	0.96
MnO	0.123	0.033	0.20	0.028	0.13	0.1
MgO	56.2	56.3	38.6	38.9	38.7	39.5
NiO	0.003	0.001	0.005	0.003	NA	0.034
CaO	0.167	0.122	0.28	0.088	0.32	0.4
Na ₂ O	0.0	0.0	0.021	0.107	0.09	0.05
Sum	100.055	99.685	99.750	98.881	100.28	100.0
Cations/24 oxygens						
P	0.0006	0.0005	0.0006	0.0006	—	—
Si	5.9769	5.9769	7.9496	7.9898	7.9762	7.864
Ti	0.0017	0.0026	0.0080	0.0008	—	0.004
Al	0.0085	0.0129	0.0620	0.0173	0.0431	0.052
Cr ³⁺	0.0257	0.0072	0.0341	0.0026	—	0.040
Fe ²⁺	0.1134	0.1019	0.1072	0.0919	0.0985	0.108
Mn	0.0147	0.0040	0.0229	0.0032	0.0149	0.012
Mg	11.8362	11.8853	7.7652	7.8652	7.8114	7.948
Ni	0.0003	0.0001	0.0005	0.0003	—	0.004
Ca	0.0253	0.0185	0.0405	0.0128	0.0464	0.056
Na	0.000	0.000	0.0055	0.0281	0.0236	0.012

Columns 1–6, electron microprobe analyses (iron given as FeO): 1, orange luminescing olivine, mean of four analyses. 2, Blue luminescing olivine, mean of three analyses. 3, Red luminescing clinostite, mean of seven analyses. 4, Blue luminescing clinostite, mean of three analyses. 5, Mean of 27 clinostite analyses from ref. 2. 6, Mean of analyses of clinostite from ref. 1. 6, Wet chemical analysis of bulk pyroxene separate⁵; FeO 0.93, Fe₂O₃ 0.03 wt %. NA, not applicable.

higher concentrations of Ti, Al, Cr, Mn and Ca, whereas blue ones have higher Na. Orange olivines contain more Cr and Mn than blue ones.

The method of assignment of valence states to minor elements in the pyroxenes has been detailed elsewhere³. There is insufficient Al and Cr³⁺ in blue clinostite to balance the observed Na, and some Fe must be assigned to the trivalent state to maintain charge balance if a defect-free MTO₃ formula is assumed. For red clinostites, it was again necessary to assume that some Fe³⁺ is present to obtain a cation: oxygen ratio fitting the theoretical value of 2:3. For clinopyroxene analyses in columns 3 and 4 of Table 1, it is necessary to assume that 38 and 8%, respectively, of total Fe is present as Fe³⁺. However, a wet chemical analysis of a pyroxene separate from Kota-Kota⁵ showed only 0.03 wt % Fe₂O₃ against 0.93 wt % FeO. The trivalent Fe was attributed to terrestrial oxidation, and the significance of the recalculation of the electron microprobe analyses is unclear. If some or all of the Cr were divalent, even more Fe³⁺ would be required.

The observed textures and bimodal chemical compositions of the olivines and clinostites rule out formation of these meteorites by direct condensation from the solar nebula. Mechanical aggregation of material from at least two source regions is needed to explain the blue and red varieties, and the

Table 2 Observed compositional variations

	Kota-Kota olivine orange	Kota-Kota pyroxene red	Indarch olivine orange	Indarch pyroxene red
P ₂ O ₅	0.003–0.008	0.000–0.010	0.000–0.007	0.000–0.017
TiO ₂	0.000–0.022	0.025–0.120	0.000–0.085	0.018–0.077
Al ₂ O ₃	0.016–0.100	0.151–0.451	0.033–0.119	0.119–0.390
Cr ₂ O ₃	0.115–0.290	0.126–0.568	0.108–0.322	0.117–0.681
MnO	0.090–0.188	0.091–0.418	0.093–0.209	0.086–0.513
NiO	0.000–0.005	0.000–0.010	0.000–0.005	0.000–0.015
CaO	0.090–0.252	0.124–0.472	0.116–0.272	0.123–0.419
Na ₂ O	0.000	0.002–0.093	0.000	0.000–0.114
	blue	blue	blue	blue
P ₂ O ₅	0.002–0.009	0.001–0.009	0.000–0.008	0.000–0.017
TiO ₂	0.002–0.031	0.001–0.012	0.000–0.047	0.000–0.014
Al ₂ O ₃	0.041–0.116	0.015–0.195	0.034–0.141	0.012–0.172
Cr ₂ O ₃	0.029–0.095	0.002–0.107	0.020–0.087	0.005–0.098
MnO	0.014–0.046	0.000–0.090	0.010–0.062	0.001–0.086
NiO	0.001–0.004	0.000–0.013	0.000–0.006	0.000–0.006
CaO	0.082–0.188	0.015–0.157	0.094–0.190	0.015–0.119
Na ₂ O	0.000	0.040–0.248	0.000	0.049–0.250

Electron microprobe analyses, wt %. Accuracy is 2–5% of amount present with detection level near 0.005 (2 σ).

observed textures indicate a complex sequence of events. Enclosure of rounded olivine grains by pyroxene grains suggests that olivine formed earlier than clinoenstatite in agreement with condensation calculations for a solar nebula with either the solar C/O ratio of 0.57 (ref. 6) or the higher ratio of ~1.0 postulated⁷ to explain enstatite chondrites.

Clinoenstatite melts incongruently at 1,557°C to form forsterite plus liquid⁸. Because the forsterite bodies are rounded rather than euhedral, it is unlikely that the forsterite observed in Kota-Kota was so formed. In addition, both orange and blue olivine occur surrounded by red or blue pyroxene. Because forsterite melts at a much higher temperature (1,850°C) than enstatite, pre-existing forsterite might have survived the melting of clinoenstatite.

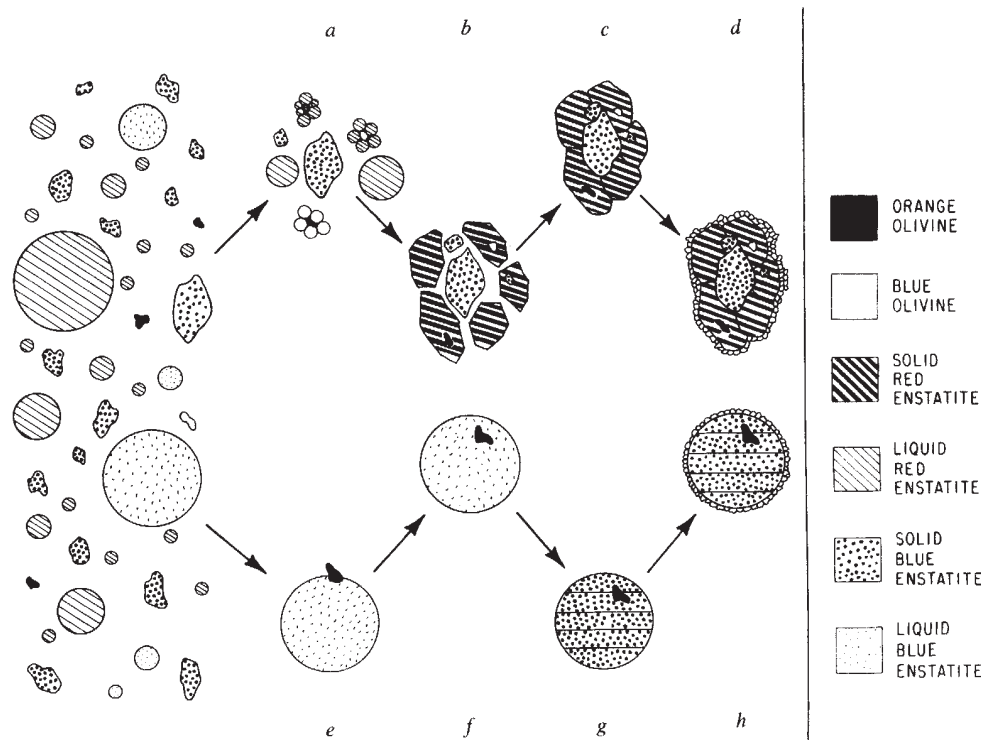
The simplest scheme to be considered for the Kota-Kota meteorite involves (1) progressive condensation of two separate

regions of the solar nebula to give minor forsterite and dominant clinoenstatite, and (2) mechanical mixing of the condensates. Because this scheme is insufficient to explain the textural properties of the chondrules, some other process or processes must be involved. In view of evidence^{9–11} for accretion of planets through planetesimals, generation of chondrules during collision of planetesimals is considered¹².

The similar ranges of clinoenstatite compositions in Table 2 indicate that the Indarch and Kota-Kota meteorites may have come from the same parent body. Orange olivine and red clinoenstatite are assumed to derive from the same condensation region because of comparable Cr₂O₃ and MnO contents, and similarly the blue olivine and blue clinoenstatite are paired. Enclosure by red clinoenstatite of blue clinoenstatite, orange olivine and blue olivine rules out mechanical mixing of crystals that condensed in equilibrium conditions in two source regions before planetesimal collision. To explain the textural relations it seems that at least some red clinoenstatite must have crystallized after the other three minerals. Because there is a sharp discontinuity in luminescence and chemistry between adjacent grains of porphyritic chondrules (Fig. 1), it seems that an enclosing grain grew without reaction with an enclosed grain. Figure 2a–d shows how the porphyritic chondrules may have formed, partly by mechanical aggregation of pre-existing crystals and partly by growth of red clinoenstatite around pre-existing crystals. A liquid stage for at least some red grains seems necessary to explain the inclusion of blue clinoenstatite and orange and blue olivines, and it may prove necessary to introduce additional steps of crystallization and mechanical disaggregation. Interstices between aggregated crystals are assumed to be filled by dust and vapour deposits.

Barred, radiating and fine-grained chondrules are assumed to have derived from liquids formed during planetesimal collision. Because these droplet chondrules are composed predominantly of either red or blue clinoenstatite with only minor amounts of interstitial phases, it is necessary to postulate formation of two separate liquids with compositions near those of the red or blue clinoenstatites. Such liquid droplets could then incorporate minor amounts of pre-existing solid material released by mechanical disaggregation of the colliding planetesimals (Fig. 2e–h). The impact melts are assumed to have been dispersed by high velocity jets¹³ produced during collision. Such jets must

Fig. 2 Schematic representation of chondrule formation from a gas-dust-liquid droplet cloud. Angular shapes are crystals; circles are liquid. a–d shows the formation of a porphyritic chondrule: a, some blue enstatite, blue olivine and orange olivine becomes surrounded by red liquid; b, red liquid forms crystals enclosing blue enstatite, blue olivine and orange olivine; c, crystals become aggregated and a porphyritic chondrule similar to that in Fig. 1 is formed; d, chondrule becomes surrounded by a shell of fine-grained dust. e–h shows the formation of a barred chondrule with an olivine inclusion: e, blue liquid droplet comes in contact with an orange olivine; f, olivine is trapped inside blue liquid droplet; g, liquid crystallizes and assumes a typical barred appearance; h, chondrule becomes surrounded by a shell of fine-grained material.



have dispersed the molten material quickly enough to prohibit mixing between the two liquids.

In such a scheme, the massive debris from the collision is assumed to coalesce into a new planetesimal in part of which melting and high-grade metamorphism would occur¹². Smaller fragments would remain in orbit, and later fall onto the surface. Dodd¹⁴ has given evidence for aerodynamic sorting of chondrules in ordinary chondrites⁴. He argued in favour of continuous entry of particles into an atmosphere because gas drag in a stagnant atmosphere would produce a graded deposit on the surface of a growing planetesimal. However, shock waves produced in a gas-dust cloud by incoming bodies from colliding planetesimals may have caused material to settle rapidly into an ungraded deposit¹².

Although details still need to be worked out, it seems certain

that enstatite chondrites are not direct products of the solar nebula. In spite of the complexity of the mineralogy and petrography, we need not abandon the well-documented mixing models for the bulk compositions of unequibrated enstatite^{15,16} chondrites, at least as a first approximation. Provided that little material was lost to the solar nebula during collision, the bulk chemistry of chondrites should be explicable in terms of direct condensates that entered the original planetesimals. Although the present scheme applies specifically to enstatite chondrites, the general features may also apply to the unequibrated ordinary chondrites^{17,18}.

We thank R. T. Dodd for helpful criticism, E. Olsen and K. Keil for specimens, NASA for grants NSG-7550 and 14-001-171, and I. M. Steele, R. Draus, I. Baltuska and O. Draughn for technical assistance.

Received 30 July 1980; accepted 8 January 1981.

1. Keil, K. *J. geophys. Res.* **73**, 6945 (1968).
2. Mason, B. *Geochim. cosmochim. Acta* **30**, 23 (1966).
3. Leitch, C. A. & Smith, J. V. *Nature* **283**, 60 (1980).
4. Allen, J. S., Nozette, S. & Wilkening, L. L. *Lunar planet. Sci.* **10**, 27 (1979).
5. Binns, R. A. *Miner. Mag.* **37**, 649 (1970).
6. Grossman, L. & Larimer, J. W. *Rev. Geophys. Space Phys.* **12**, 71 (1974).
7. Larimer, J. W. & Bartholomay, M. *Geochim. cosmochim. Acta* **43**, 1455 (1979).
8. Bowen, N. L. & Andersen, O. *Am. J. Sci.* 4th Ser. **37**, 488 (1914).
9. Safronov, V. S. *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets* (Israel Program for Scientific Translations, Jerusalem, 1972).
10. Wetherill, G. W. *Proc. 7th Lunar Sci. Conf.* 3245 (1976).
11. Smith, J. V. *Miner. Mag.* **43**, 1 (1979).
12. Smith, J. V. *Lunar planet. Sci.* (in the press).
13. Kieffer, S. *Science* **189**, 333 (1975).
14. Dodd, R. T. *Earth planet. Sci. Lett.* **30**, 281 (1976).
15. Larimer, J. W. & Anders, E. *Geochim. cosmochim. Acta* **31**, 1239 (1967).
16. Ikramuddin, M., Binns, C. M. & Lipschutz, M. E. *Geochim. cosmochim. Acta* **40**, 133 (1976).
17. Hutchison, R., Bevan, A. W. R., Agrell, S. O. & Ashworth, J. R. *Nature* **280**, 116 (1979).
18. Bevan, A. W. R. & Axon, H. J. *Earth planet. Sci. Lett.* **47**, 353 (1980).

Porous anodic film formation on aluminium

G. E. Thompson & G. C. Wood

Corrosion and Protection Centre, University of Manchester Institute of Science and Technology, Manchester M60 1QD, UK

The development of porous anodic films on aluminium in the major forming acids has been established¹⁻⁵. However, little information has been provided to explain the markedly different rates of steady-state film growth for anodizing in constant voltage conditions in these acids, with the anodic film cell parameters dependent on the cell voltage¹. The regular cellular morphology of the steady-state films and the substructure developed in the forming acids are revealed in Fig. 1. The characteristic parallel-sided pores, at the centres of approximately hexagonal cells, pass normal to the macroscopic metal surface but are separated from it by the barrier layer. Here, from considerations of the film material substructure and the distribution of species within the film derived from the forming acid, it is postulated that the likely variations in the local field strengths across the barrier layers can explain the steady-state anodizing behaviour observed in the different acids in constant voltage conditions.

Direct transmission electron microscopy of ion-beam thinned steady-state porous anodic films, formed originally in phosphoric or oxalic acid, has shown the presence of cell boundary bands separated from the electrolyte in the pores by differently textured film material⁶ (Fig. 1). Microanalysis of the regions comprising the cell material of the film formed in phosphoric acid suggests that the cell-boundary band is composed of virtually pure alumina, whereas phosphorus is detected in the adjacent differently textured material⁷. For films formed in chromic acid, no significant chromium is detected in the cell material⁵, which appears similar in nature to the cell boundary band material observed for films formed in phosphoric acid, that is, relatively pure alumina. The highly acid anion contaminated films formed in sulphuric acid also show a region of relatively pure alumina which is not resolved readily in the electron microscope for the films formed at the usual low voltages (R. Ozcan and our unpublished work). From such knowledge it is possible to depict the barrier layers of the films formed in the major anodizing acids in terms of the ratio of the thicknesses of

relatively pure alumina region to acid anion-contaminated material, which seems to increase in the order: sulphuric acid (0.05) > oxalic acid (0.1) > phosphoric acid (0.5) > chromic acid (∞).

The relative extents of the different regions present in the barrier layers of the films formed at constant voltages in the usual anodizing conditions in the above acids are illustrated in Fig. 1, which shows that the films form at relatively similar nm V^{-1} ratios. The relatively pure alumina is drawn as a solid region as it is thought that this more compact film grows by ionic migration to form a glassy or anion-free microcrystalline structure, whereas the outer microcrystalline acid anion-contaminated film material is deposited under the field from initially colloidal alumina⁸. Different extents of deposition explain the varying current efficiencies observed in the corresponding acids.

There is a direct link between the above findings and the rates of formation of the films in the different anodizing acids, at constant voltage. For films formed at 20 V, typical rates of formation increase in the order: sulphuric acid > oxalic acid > phosphoric acid > chromic acid.

For example, the rate of formation of the film in sulphuric acid, with the thinnest relatively pure alumina region, is much greater than in chromic acid, where films composed of relatively pure alumina are formed. Furthermore, the current density (i) is related to the potential drop (V) across the barrier layer by the high field conduction theory:

$$i = A \exp(BV/d)$$

where A and B are constants, for a given temperature, and d is the overall barrier layer thickness. In the above equation V/d represents the effective average field across the barrier layer which is approximately constant for the barrier layers of the different films because the measured nm V^{-1} ratios are similar. However, if the true field across the relatively pure alumina region of the barrier layer is in the order: sulphuric acid > oxalic acid > phosphoric acid > chromic acid, because this represents the order of increasing thickness of the observed cell boundary bands, then the rates of formation of the films in the different acids will also vary in the order given above if solid state ionic migration across the inner layer is the rate-determining step. The passage of ions through the outer regions of the contaminated film is thought to proceed relatively easily. Thus, during anodizing in the major anodizing acids, the field strength is thought to be greater across the inner, compact relatively pure

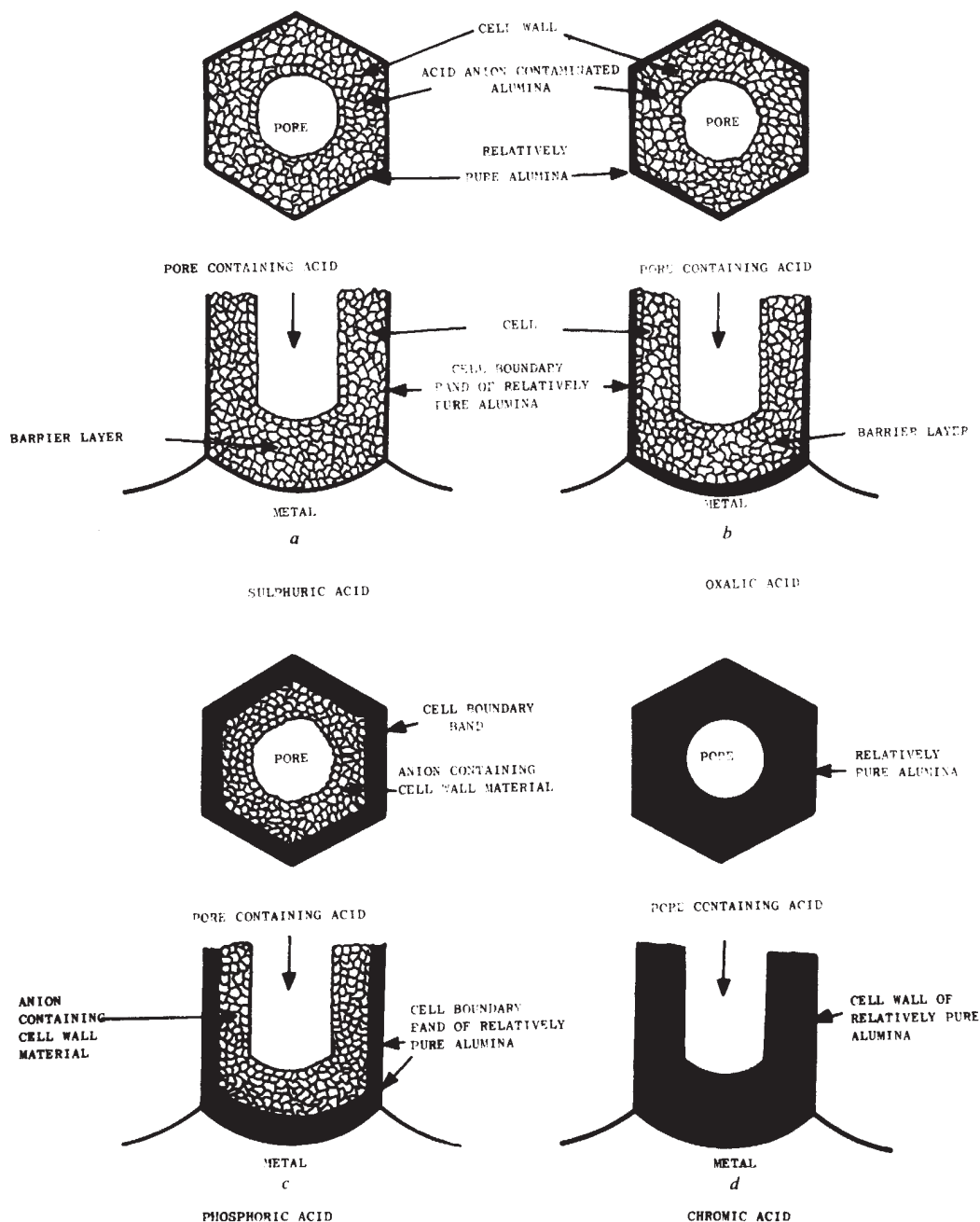


Fig. 1 The plan and sectional views of a pore and adjacent cell for films formed in each of the major acids: *a*, sulphuric acid; *b*, oxalic acid; *c*, phosphoric acid; *d*, chromic acid. The cell comprises acid anion-contaminated film material adjacent to the pore and relatively pure alumina remote from it: □ acid anion-contaminated film material; ■, relatively pure alumina.

alumina region and lower across the acid anion-contaminated alumina film, which probably varies from solid in contact with the cell boundary bands to gel-like material at the base of the pores. Microporosity or relatively easy passage of ions through intercrystallite pathways will reduce the effective field further. Figure 2 shows how the total voltage drop across the barrier layer is similar for the highly contaminated film formed in sulphuric acid (Fig. 2*a*) and the acid anion-free film formed in chromic acid (Fig. 2*d*) where the cell boundary band to contaminated film material ratios approach 0 and ∞ , respectively. The current density is higher in the former case because of the higher voltage drop across the pure alumina region adjacent to the metal interface. In addition, anodizing in chromic acid proceeds at relatively low current efficiency. In phosphoric and oxalic acids there are two zones. The voltage drop and field across the pure cell boundary bands are high but low in the outer acid anion-contaminated film material (Fig. 2*b, c*).

The variation in voltage drop across the adjacent film regions comprising the barrier layer is important and has been considered by Dell'Oca and Young⁹ in an investigation of the

different layered barrier film formed on tantalum in phosphoric acid. They suggested that a high field operates in the contaminated layer adjacent to the solution where phosphate reduces the permittivity and decreases ionic mobility. The case for a dependence of field strength on film thickness (for thin films) has been put forward by Dignam¹⁰. Here the different field strengths have been based on the likely mechanisms of film growth during steady-state porous film formation involving solid-state ionic migration across the relatively pure alumina region and deposition of the outer layer from a hydrous metal oxide solution. Direct observations of the film materials, revealed by ultramicrotomy of both barrier and porous-type anodic films, show the sections to be relatively uniform and featureless^{2,3,8,9,11}, only revealing telling changes after exposure to the electron beam¹⁰. This is thought to represent 'drying out' and/or ageing of the film material after anodizing, and it is only during anodizing that the variation in field strengths across the barrier layer is manifested. As the nm V⁻¹ ratios for the different films have been determined after anodizing in the absence of the field it is not surprising that they show relatively similar values.

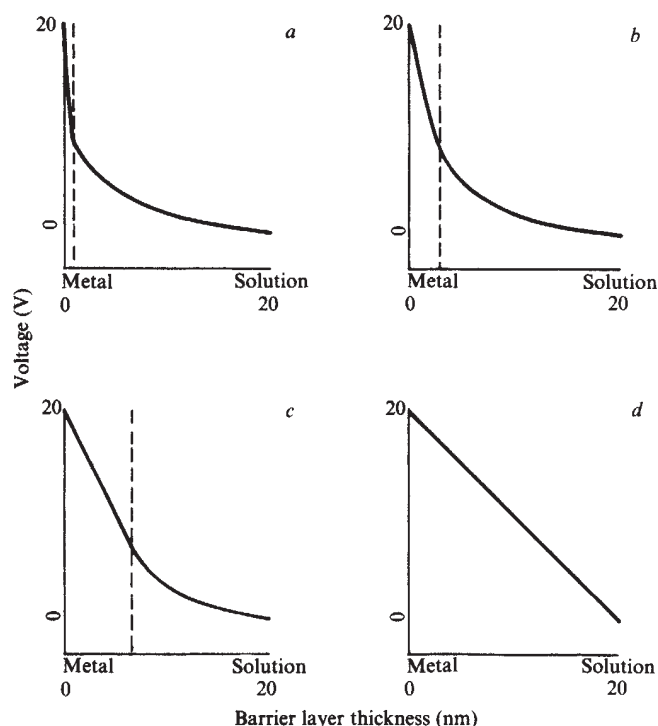


Fig. 2 Distribution of the voltage drop and field (the slope of the voltage-distance plot) across typical barrier layers of porous anodic films formed in each of the major acids: *a*, sulphuric acid; *b*, oxalic acid; *c*, phosphoric acid; *d*, chromic acid. The voltage drop is greatest and linear across the inner compact relatively pure alumina region and lower across the outer acid anion contaminated film material, where the voltage decreases progressively towards the film-solution interface. Consequently, the highest field is across the least extensive relatively pure alumina region, evident for the films formed in sulphuric acid, decreasing for the films formed in oxalic and phosphoric acids respectively, and lowest for the most extensive pure alumina region present in the films formed in chromic acid. The rates of formation of the corresponding films decreases in the same order.

The nature of the incorporated acid anion will probably strongly influence the size distribution of the deposited and contaminated microcrystalline alumina¹², and its hydrogen bonding characteristics, which affect the effective compaction of this material.

Thus, the steady-state anodizing behaviour of porous anodic films formed in the major anodizing acids can be related to the distribution of the acid anions within the barrier layers and the true field strengths across the relatively pure alumina regions. The effect of the acid anion on thermally enhanced field-assisted dissolution, or other processes whereby the relatively pure alumina region transforms to the acid anion-incorporated material and its influence on anodizing at constant current density, will be discussed elsewhere.

Received 18 November 1980; accepted 9 January 1981.

1. O'Sullivan, J. P. & Wood, G. C. *Proc. R. Soc. A* **317**, 511-543 (1979).
2. Thompson, G. E., Furneaux, R. C., Goode, J. S. & Wood, G. C. *Trans. Inst. Metal Finishing* **56**, 159-167 (1978).
3. Thompson, G. E., Furneaux, R. C. & Wood, G. C. *Trans. Inst. Metal Finishing* **55**, 117-128 (1977).
4. Bailey, G. & Wood, G. C. *Trans. Inst. Metal Finishing* **52**, 187-199 (1975).
5. Thompson, G. E., Wood, G. C. & Hutchings, R. *Trans. Inst. Metal Finishing* **58**, 21-25 (1980).
6. Thompson, G. E., Furneaux, R. C. & Wood, G. C. *Corros. Sci.* **18**, 481-498 (1978).
7. Thompson, G. E., Furneaux, R. C., Wood, G. C. & Hutchings, R. *J. electrochem. Soc.* **125**, 1480-1482 (1978).
8. Thompson, G. E., Furneaux, R. C., Wood, G. C., Richardson, J. A. & Goode, J. S. *Nature* **272**, 433-435 (1978).
9. Dell'Oca, C. J. & Young, L. *J. electrochem. Soc.* **117**, 1548-1551 (1970).
10. Dignam, M. J. *J. electrochem. Soc.* **126**, 2188-2195 (1979).
11. Furneaux, R. C., Thompson, G. E. & Wood, G. C. *Corros. Sci.* **18**, 853-881 (1978).
12. Matyevic, E., Bell, A., Brace, R. & McFadyen, P. *J. electrochem. Soc.* **120**, 893-899 (1973).

Tasmanian tree rings belie suggested anthropogenic $^{13}\text{C}/^{12}\text{C}$ trends

R. J. Francey

CSIRO Division of Atmospheric Physics, Aspendale, Victoria, Australia 3195

Published measurements in tree rings¹⁻¹⁰ suggest a persistent decrease with time of the ratio $^{13}\text{C}/^{12}\text{C}$, particularly in the period 1860-1940, and the release of C^{13} -deficient CO_2 from both fossil fuel combustion and forest clearing have often been invoked to explain the decreases. Freyer¹¹ concludes that "the existing measurements leave no doubt that the ^{13}C content in the cellulose fraction of modern wood was indeed decreased significantly". Data presented here for two tree species from sites with contrasting ecologies on the remote and exposed Tasmanian west coast do not, however, support claims that these trends, which have been observed in (mostly) Northern Hemisphere trees, are the global response of the atmosphere to agricultural and industrial activities.

Freyer's statement requires qualification. There is evidence^{2,4,7-9} of a levelling or an increase in $^{13}\text{C}/^{12}\text{C}$ ratios after 1940, moreover the period before 1860 is poorly sampled, thus the statistical significance of a modern trend over several decades is difficult to assess. The Freyer's conclusion is based

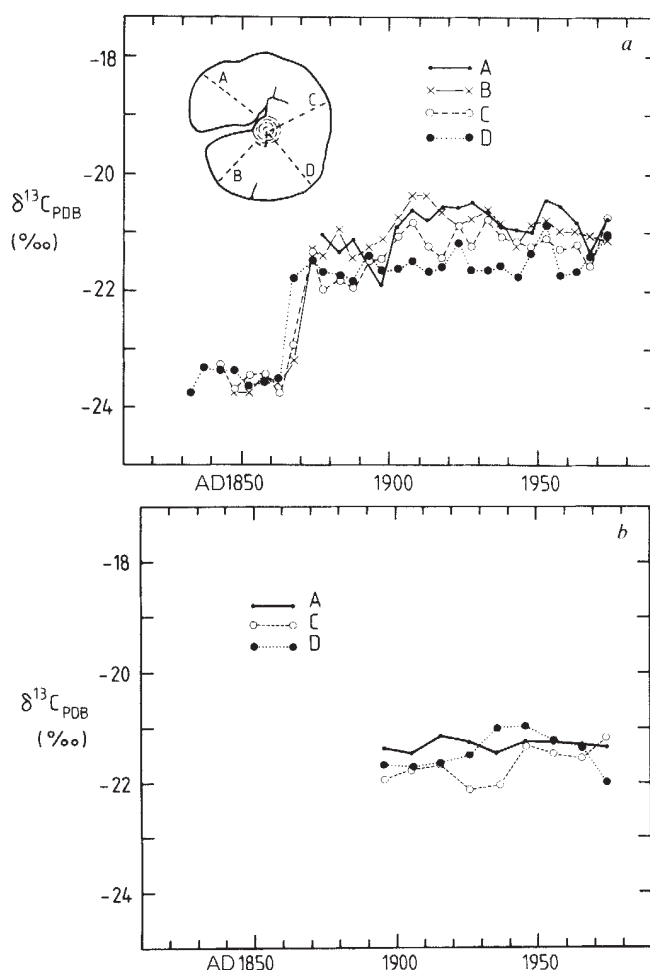


Fig. 1 *a*, $^{13}\text{C}/^{12}\text{C}$ analyses on 5-yr blocks from four radii of tree JR39. Radii were cut from a disk as shown in the inset. *b*, $^{13}\text{C}/^{12}\text{C}$ analyses on 10-yr blocks from three radii of tree BH858.

Table 1 Tasmanian trees analysed for $\delta^{13}\text{C}$

Tree	Species	Site description		Ecology	Approximated germination (AD)	Period analysed (AD)
		Altitude (m)	Distance to ocean* (km)			
SA†	<i>Athrotaxis selaginoides</i>	880	90	Rain forest	<1700	1841–1974
S3†	<i>Athrotaxis selaginoides</i>	880	90	Rain forest	<1700	1893–1974
SC15‡	<i>Athrotaxis selaginoides</i>	880	90	Rain forest	1090	1540–1956
SC19‡	<i>Athrotaxis selaginoides</i>	880	90	Rain forest	1677	1743–1952
MR3.4‡	<i>Athrotaxis selaginoides</i>	1,100	30	Exposed mountain top	<1450	1500–1975
BH858§	<i>Phyllocladus aspleniifolius</i>	40	15	Small island	1880	1885–1958
JR39§	<i>Phyllocladus aspleniifolius</i>	60	35	Rain forest	1829	1830–1977

* In a southwesterly direction.

Dendrochronology from: † D. Adamson (Macquarie University); ‡ J. Ogden (University of Auckland, formerly Australian National University);

§ T. Bird (CSIRO Division of Forestry).

largely on trees of mid-latitude Northern Hemisphere origin, close (in a global sense) to major anthropogenic sources of CO_2 and associated pollutants¹², embedded in large natural seasonal variations in atmospheric CO_2 , and remote from major exchange reservoirs in the southern oceans.

The further step of relating a tree ring derived $^{13}\text{C}/^{12}\text{C}$ trend to the global CO_2 budget presents its own problems⁹, including (1) ignorance of the physical factors influencing $^{13}\text{C}/^{12}\text{C}$ in cellulose, sometimes creating large differences in cellulose of similar age from different locations within a tree, (2) the influence of local environmental factors, and (3) the relative insensitivity of $^{13}\text{C}/^{12}\text{C}$ in the atmosphere–biosphere–ocean system to gradual changes in the biosphere.

The forests of the Tasmanian west coast offer many advantages in the search for a global perspective on the carbon dioxide history of the atmosphere. The area is predominantly wilderness, remote from large scale industry and exposed to the almost circumpolar ocean fetches of the 'roaring forties'. Seasonal variations in the atmospheric CO_2 concentration are less than a tenth of those at high northern latitudes¹³. The native flora includes several species suitable for ring-width dating^{14–16} and living trees of great age, in stable ecosystems, are available.

Results are now available from the first seven trees analysed, whole wood analyses from three of which have been reported elsewhere^{17,18} and the temperature correlations reported there, while still evident at reduced magnitude, do not conflict with the cellulose results reported here. In what follows, comparisons will be made with measurements on cellulose^{1,3,4,8,9} (though similar conclusions would come from consideration of whole wood analyses^{2,4,7}).

The trees are referred to by their collector's designation, and relevant properties are summarized in Table 1. Full details of the trees and the analyses techniques will be given elsewhere¹⁹. The analysis repeatability on a homogeneous wood sample is typically 0.06‰.

On two of the trees, JR39 and BH858, comparisons have been made of $^{13}\text{C}/^{12}\text{C}$ variations along different radii. Analyses on 5-yr blocks of four radii for JR39 are shown in Fig. 1a, the inset shows the radii cut from a disk.

This record shows sharp discontinuity in $\delta^{13}\text{C}_{\text{PDB}}$ occurring at around 1869, where scarring suggests fire damage. After 1869 there is a corresponding sharp increase in ring width. The $\delta^{13}\text{C}_{\text{PDB}}$ and ring width increases are probably the result of sudden exposure of a sub-canopy sapling to free atmosphere, removing limiting factors related to competition, and/or markedly reducing the uptake of ^{13}C depleted respired carbon dioxide²⁰. It suggests that variations in $\delta^{13}\text{C}$ after this event were not seriously influenced by these factors.

While $\delta^{13}\text{C}_{\text{PDB}}$ varies by up to 1‰ in different radii, these differences are generally maintained throughout the adult life of the tree. In the context of possible large monotonic trends¹¹, this is taken as evidence (with some caution⁹) that trends representative of the whole disk can be inferred from one radius.

This application when only one radius is available is supported by Fig. 1b, showing three cores from tree BH858. The data are analysed in 10-yr blocks. Radius A is a core from a south-west direction, 1.09 m from the base; C is a north-east core from 0.58 m and D a south-east core from 0.81 m. Although a relatively young rainforest tree, there is no juvenile effect. It may be relevant that it is located on a small (~100 m diameter)

Fig. 2 $^{13}\text{C}/^{12}\text{C}$ analyses on all seven trees. Averages of the different radii are used for JR39, BH858. (Dashed lines indicate values corresponding to periods of marked ring width discrepancies.)

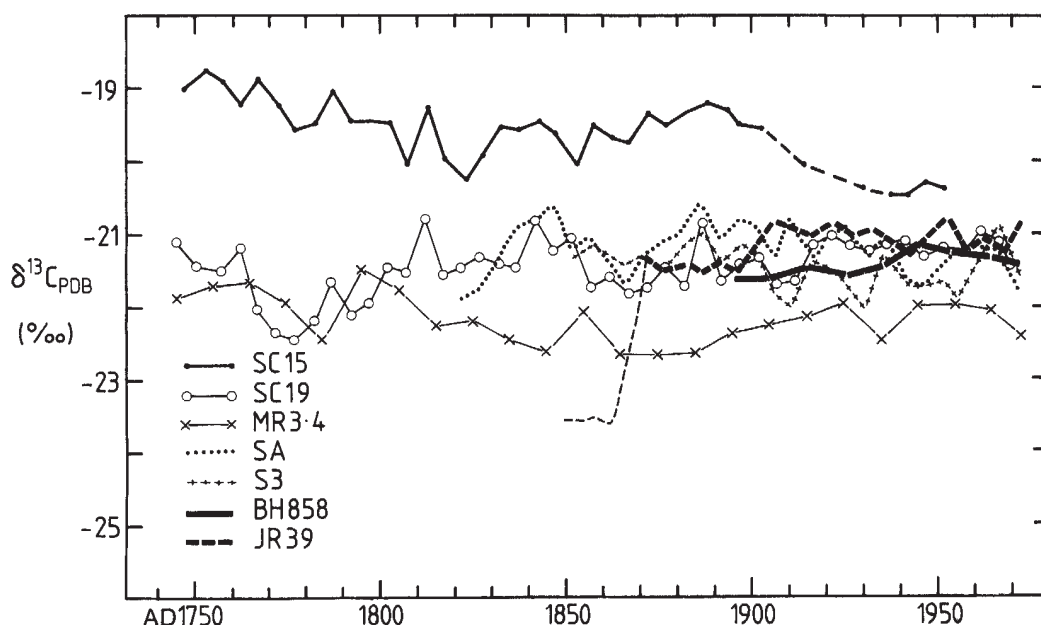


Table 2 Comparison of $\delta^{13}\text{C}$ trends since 1950

Medium	Period of data	Regression coefficient (% yr ⁻¹)	Ref.
Atmosphere	1956–1978	-0.030	21, 22
Cellulose	1950–1972	-0.018 ± 0.008	3
Cellulose	1950–1972	+0.042 ± 0.011	9
Cellulose	1951–1974	-0.005 ± 0.008	This work (6 tree averages)

exposed island in Bathurst Harbour (separated by 15 km of water and grassed hills from the open ocean), and flushing of respired sub-canopy CO_2 should be much more efficient than in a land-locked, heavily vegetated area.

Figure 2 shows an average curve for each of these trees, the remaining five curves referring to a single radius per tree. Four of these trees (SC15, SC19, SA, S3) are from a rainforest on a mountain ridge well exposed to the south-west, but 90 km from the coast. Three had germination dates estimated at about 1700 (although samples relating to a briefer period were provided for analysis) and were probably well exposed to free atmosphere from 1800. SC15, however, is much older with an estimated germination about 1090. Ring width data indicate a severe setback in growth from 1907 to 1917. In blocks after 1907, $\delta^{13}\text{C}_{\text{PDB}}$ values are on average about 1‰ more negative than immediately before 1907. At the time of sampling SC15 was lying on the ground and the tree may have fallen (had achieved a sub-canopy status) as early as 1907.

There is no possibility of canopy influences over the past several hundred years for MR3.4: estimated germination precedes 1450, and the high altitude, stunted 'krumholtz' form is evidence of its isolation and extreme exposure.

Figure 3 shows the combined data although the first 30 yr of JR39 analyses and the last 50 yr of SC15, (indicated by dashed lines in Fig. 2) for which independent ring width evidence suggest sub-canopy influences, are excluded. SC15, SC19, MR3.4 have common records between 1740 and 1905; and in Fig. 3a averages (in 10-yr blocks) of their deviations from the 1740–1900 mean are shown.

Six trees have common records between 1890 and 1970, and five of these extend back to 1870. Figure 3b shows averages (in 10-yr blocks) of their deviations from the 1890–1970 means (using five trees before 1890 and six trees thereafter).

In the context of global changes in atmospheric $\delta^{13}\text{C}$, the average of the six trees in Fig. 3b exhibits a striking lack of systematic behaviour—the data for a 10-yr interval never deviate significantly from the 1870–1970 mean. This contrasts with

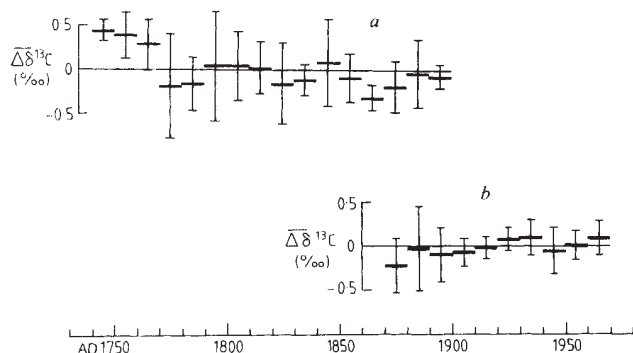


Fig. 3 Combined $^{13}\text{C}/^{12}\text{C}$ data from 10-yr blocks. *a*, The average and standard deviation of deviations from the 1740 to 1900 mean for SC15, SC19, MR3.4. *b*, The average and standard deviation of deviations from the 1890–1970 means for SC19, MR3.4, JR39, BH858, SA and S3; for the period 1870–1890 only BH858 is not represented.

the steady decrease of about 0.5‰ (from 1870 to 1960) of Freyer¹¹.

A large decrease in $\delta^{13}\text{C}$ of almost 1‰ between 1840 and 1920 of Freyer¹¹, frequently interpreted as a global atmospheric response to a large net transfer of CO_2 from the biosphere^{3,6,8}, is not supported by the results in Fig. 3a. Clearly Fig. 3 and Freyer's results, for example, cannot both represent global atmospheric $^{13}\text{C}/^{12}\text{C}$ behaviour.

All Northern Hemisphere cellulose data are not incompatible with the Tasmanian results; over the same 1870–1970 period of Fig. 3b, the mean curve of Tans and Mook⁹ (and the whole wood data of Farmer⁷ and Galimov⁴) show a similar absence of a monotonic trend.

The data of Keeling *et al.*^{21,22} provide a useful comparison with the above results. They report annual average values of $\delta^{13}\text{C}$ and μ (the total CO_2 concentration);

$$1956 \mu = 314.1 \text{ p.p.m.} \quad \delta^{13}\text{C} = -6.69\text{‰}$$

$$1978 \mu = 334.2 \text{ p.p.m.} \quad \delta^{13}\text{C} = -7.34\text{‰}$$

representing a $\delta^{13}\text{C}$ change of -0.6‰ over 22 yr, or -0.030‰ yr⁻¹ for a linear trend.

Table 2 shows the available individual measurements of $\delta^{13}\text{C}$ in cellulose from 1950 linearly regressed against time. The anticipated atmospheric $\delta^{13}\text{C}$ response to fossil fuel combustion alone⁹ suggests that a linear approximation is reasonable over this period, and that the Table 2 regression coefficients (plus standard errors) are directly comparable.

Assuming the results obtained by Keeling *et al.*²¹ represent an ubiquitous $\delta^{13}\text{C}$ trend in the free atmosphere (supported by their comparison of North American and South Pole $\delta^{13}\text{C}$ measurements in 1978), then only the Rebello and Wagener tree can have responded in a direct manner.

The differences between $^{13}\text{C}/^{12}\text{C}$ trends in trees from different regions, coupled with the differences between most $^{13}\text{C}/^{12}\text{C}$ trends in tree-rings representing the past 20 yr and the direct atmospheric measurement, seriously question the interpretation of any tree record in terms of global atmospheric behaviour, and in particular the implicit assumption of constant fractionation in the atmosphere-biosphere exchange. An explanation involving large scale modification of atmospheric $^{13}\text{C}/^{12}\text{C}$ ratios as a result of plant respiration²³, apparently conflicts with well known rates of mixing of CO_2 in the atmosphere²⁴ during photosynthesizing hours. A solution must be sought in the mechanism of fractionation in plants.

I thank Dr Don Adamson, Dr John Ogden, Trevor Bird and George Dolezal for dated timber and detail on the individual trees, John Ogden and Drs Graeme Pearman and Paul Fraser for discussions, and Geoff Richards for the bulk of the chemical preparation.

Note added in proof: With respect to the suggestion²⁵ that isotope variations might primarily reflect ring width changes, there is a general absence of correlation between $\delta^{13}\text{C}$ and ring width in these trees, apart from exceptions mentioned in the text.

Received 28 August 1980; accepted 14 January 1981.

1. Freyer, H. D. & Wiesberg, L. *Naturwissenschaften* **60**, 517–518 (1973).
2. Farmer, J. G. & Baxter, M. S. *Nature* **247**, 273–275 (1974).
3. Rebello, A. & Wagner, K. *Proc. int. Symp. on Environmental Biogeochemistry* Vol. 1 (ed. Nriagu, J. O.) 13–23 (Ann Arbor, Michigan, 1976).
4. Galimov, E. M. *Proc. int. Symp. on Environmental Biogeochemistry* Vol. 1 (ed. Nriagu J. O.) 3–11 (Ann Arbor, Michigan, 1976).
5. Stuiver, M. *Science* **199**, 253–258 (1978).
6. Wilson, A. *Nature* **273**, 40–41 (1978).
7. Farmer, J. G. *Nature* **279**, 229–231 (1979).
8. Freyer, H. D. *Tellus* **31**, 124–137 (1979).
9. Tans, P. P. & Mook, W. G. *Tellus* **32**, 268–283 (1980).
10. Grinstead, M. J., Wilson, A. T. & Ferguson, C. W. *Earth planet. Sci. Lett.* **42**, 251–253 (1979).
11. Freyer, H. D. *WMO Project on Research and Monitoring of Atmospheric CO_2* Rep. No. 1, Appendix E (WMO Geneva, 1979).
12. Freyer, H. D. *Tellus* **31**, 308–312 (1979).
13. Pearman, G. I. *Clean Air* 20–26 (1977).
14. Ogden, J. *Aust. J. Ecol.* **3**, 339–356 (1978).
15. Ogden, J. *Tree-Ring Bull.* **38**, 1–13 (1978).
16. LaMarche, V. C. Jr., Holmes, R. L., Dunwiddie, P. W. & Drew, L. G. *Tree-ring Chronologies of the Southern Hemisphere* Vol. 4 (Laboratory of Tree-Ring Research, Arizona, 1979).

17. Pearman, G. I., Francey, R. J. & Fraser, P. J. B. *Nature* **260**, 771–773 (1976).
18. Fraser, P. J. B., Francey, R. J. & Pearman, G. I. *DSIR Bull.* **220**, 67–73 (1978).
19. Francey, R. J. *et al.* (in preparation).
20. Keeling, C. D. *Geochim. cosmochim. Acta* **24**, 299–313 (1961).
21. Keeling, C. D., Mook, W. G. & Tans, P. P. *Nature* **277**, 121–123 (1979).
22. Keeling, C. D., Bacastow, R. B. & Tans, P. P. *Geophys. Res. Lett.* **7**, 505–508 (1980).
23. Mazany, T., Lerman, J. C. & Long, A. *Nature* **287**, 432–435 (1980).
24. Keeling, C. D. *Geochim. cosmochim. Acta* **13**, 322–334 (1958).
25. Wigley, T. M. L., Gray, B. M. & Kelly, P. M. *Nature* **271**, 92–93 (1978).

Ixtoc 1 oil spill: flaking of surface mousse in the Gulf of Mexico

John S. Patton & Mark W. Rigler

Department of Microbiology, University of Georgia, Athens, Georgia 30602, USA

Paul D. Boehm & David L. Fiest

Energy Resources Co. Inc., 185 Alewife Brook Parkway, Cambridge, Massachusetts 02138, USA

The blowout at the Ixtoc 1 offshore oil rig in the Gulf of Mexico (19°24' N, 92°12' W) produced a major oil spill. An estimated 140 million gallons of oil escaped before the well was finally capped¹; this compares with a spill of 66 million gallons of oil from the tanker *Amoco Cadiz*^{2,3}. In contrast to the rapid, close-to-shore release of the tanker's cargo, the Ixtoc 1 well released oil for 9 months into the open ocean where winds and currents dispersed the floating mousse (a light brown and orange water-in-oil emulsion) which had formed at the wellhead. Two and a half months after the blowout the western boundary current in the Gulf of Mexico^{4,5} had brought mousse to the American shore. An ultra-thin sheen also formed from the oil and was visible as a smoothness of sea-surface texture relative to the slightly rougher unpolluted waters. We report here that at a distance of 750–1,000 km from the well, Ixtoc 1 mousse occurred on sheen-covered sea surface primarily as gramme-sized pancakes and milligramme-sized flakes. Relative to pancakes, flakes were more dense, depleted in aliphatic compounds and enriched in polar compounds. Both particles were covered with a brittle 5–10 µm skin. Pancakes, skin, flakes and surface sheen seem to represent the four macroscopic component parts of a pelagic weathering cycle of Ixtoc 1 mousse.

Samples were collected from three stations near the South Texas coast and from one station near the wellhead (Table 1). Generally mousse particles occurred within or very close to areas covered with sheen, at an estimated concentration, based on scale-referenced sea-surface photographs, of 0.2–2.0 g mousse per m² of sea surface. Divers estimated that 99% of the mousse occurred on the sea surface. Flakes and pancakes did not always occur in the same sheen area. Flakes were collected with a stainless-steel screen and pancakes with a dip net. Station 1 was in a flake-pancake area of at least 20 km width, with flake concentrations ranging from 92 to 158 flakes m⁻² (470–806 mg flake m⁻²). This is ~400 times above the normal background tar concentration in that area of the Gulf⁶. Both types of mousse usually occurred in the shape of flattened disks with thickness usually less than one-fifth of the width. Samples were stored in filtered seawater in the dark at 4 °C in the presence of 0.02% sodium azide⁷. The visible production of thin sheen from mousse was not observed although at one point between stations, in the middle of a hot calm day, sheen was seen to be emanating from pancakes. Large numbers of flakes or pancakes in small volumes of water showed no tendency to aggregate. There was no size continuum between pancakes and flakes; the 1–4 mm thick flakes seemed to originate from the cracked scaly crust of the pancakes (Fig. 1a) and flakes could be produced by shaking pancakes in seawater.

Microscopic views of flakes are shown in Fig. 1b–h. Figure 1b shows the penetration of light through the edges of two flakes

observed with normal transmission light microscopy. Figure 1c, d shows the surface topography of two flakes viewed by reflected epifluorescence microscopy. Figure 1c shows the cracked surface of a flake that was touched by a Pasteur pipette. The light-coloured skin is shown on edge (S) and averaged 5–10 µm in thickness. The dark material within the cracks is emulsion (see also Fig. 1g). Figure 1d shows another flake with a more heterogeneous surface. The colour of the thin skin on an individual flake ranged from light yellow to brown. Often emulsion such as that in the upper-right corner of Fig. 1d was speckled with amorphous white fluorescent particles. Figure 1e shows a freeze-etched electron micrograph of the surface of flake skin with embedded diatom tests (D) and hundreds of flattened half-moon-shaped disks. Figure 1f is a blowup of the unidentified disks that are embedded flat and on end in the skin surface. Compared with flakes from Stations 2 and 3 flakes from Station 1 were highly contaminated with embedded material. Exposed emulsion on the surface of a flake is shown in Fig. 1g. Microbial colonies (M) were uncommon and usually occurred in isolated pockets on the flake. Loosely attached microbial cells and slime, such as on the edge of the flake in Fig. 1c, may have been more common but lost during sample transport. Microbial colonies were not observed on the brittle skin. Figure 1h shows a freeze-etched view of the interior of a flake that shows a smooth textured crude oil⁸ with a heterogeneous population of emulsified water droplets. The interior of flakes contained embedded particles such as dinoflagellate and diatom tests, fish scales and unidentified amorphous material. The pancake surfaces showed the same microscopic features as the flakes.

The pancake mousse exhibited gross chemical similarity to the wellhead mousse (Table 1), although their compositions were quite different: the latter containing significant quantities of saturated hydrocarbons in the *n*-C₁₀ to *n*-C₁₅ region (Fig. 2a) and large amounts of light aromatics (Fig. 2b), whereas the pancake mousse was almost totally depleted in alkanes eluting before *n*-C₁₅ and weathered such that only small levels of two ringed aromatics remained (Fig. 2c, d).

Mousse pancakes and their associated flakes were also compositionally different, (Table 1): the flakes were depleted in saturated hydrocarbons (*f*₁) and enriched in polar (*f*₃) content relative to the pancakes. Flakes were also more dense than their companion pancakes (Table 1) which could reflect a greater volume concentration of surface-embedded material than with the pancakes. Gas chromatograms of the mousse pancakes and flakes (Fig. 2c–e) indicate subtle but significant compositional differences between the two Ixtoc oil forms. The saturated (aliphatic) fraction of the flakes is more depleted in the lower boiling *n*-alkanes and reveals an unresolved complex mixture (UCM). Quantities of residual aromatics (Fig. 2d, f), consisting mainly of alkylated phenanthrenes and dibenzothiophenes, were reduced in the flakes although compositionally the two forms were similar. Figure 3 summarizes the hydrocarbon composition of wellhead mousse, Texas mousse pancakes and Texas mousse flakes. The presence of alkylated naphthalenes in the pancakes and their absence in the flakes indicates differences in the levels of the more soluble and potentially toxic naphthalenes in the mousse.

Changes in mousse composition can be attributed to rapid evaporation and dissolution⁹ in the early weathering stages and to photooxidation and evaporation in the later stages¹⁰. Differences between flake and skin compositions, reflected in severe depletion of aromatic hydrocarbons in the latter (compare Fig. 2f with h) are probably attributable to the importance of photooxidation in the skin. The high density of skin (>1.022 g cm⁻³) (Table 1) was due to a high proportion of embedded material.

The world oceans are widely polluted with floating petroleum residues—pelagic tar^{11–15}. Approximately one-third of the pelagic tar samples from the Gulf of Mexico¹⁶ and more than half of those from the central North Atlantic¹⁷ exhibit biomodal paraffin distributions in their gas chromatograms which indicate that the residues come from tanker sludge¹⁸. The remaining

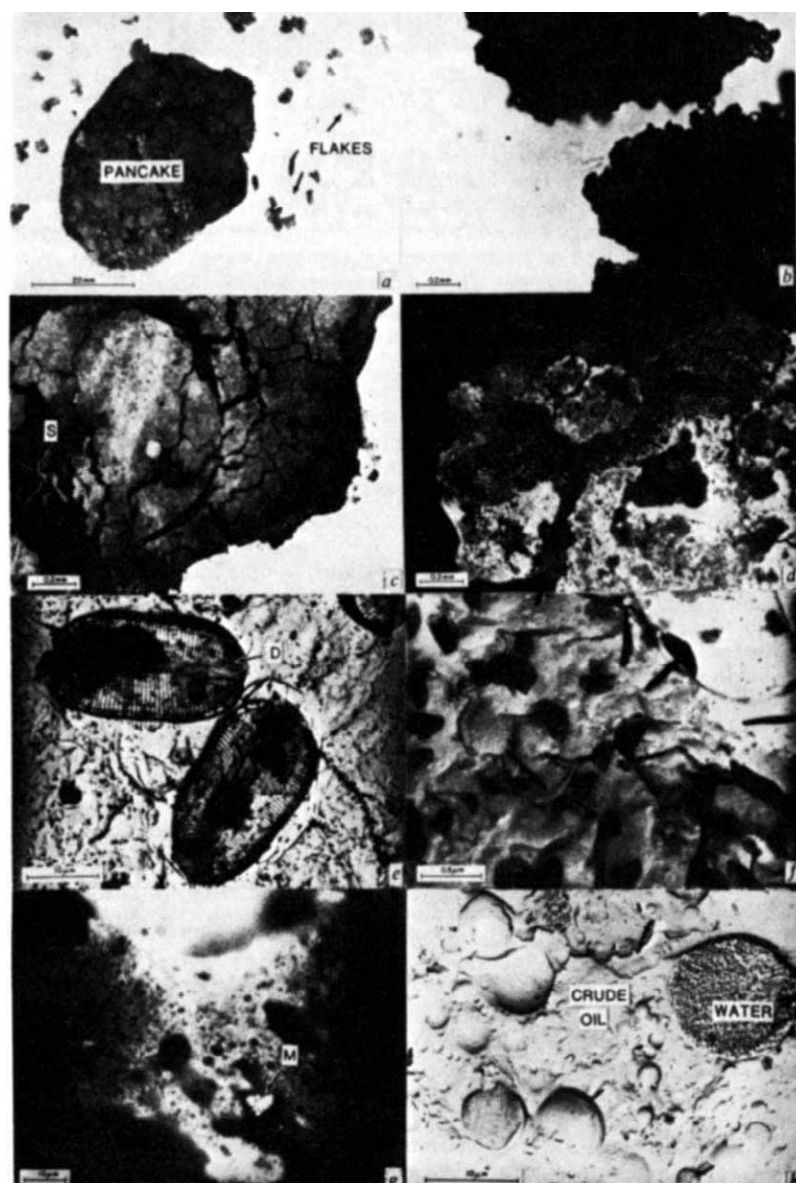


Fig. 1 A microscopic view of mousse flakes produced by weathering of Ixtoc 1 mousse pancakes in the Gulf of Mexico. *a*, A typical pancake and numerous flakes taken at Station 2 from the sea surface. *b*, Tungsten light penetration of two flakes from Station 2. *c*, Epifluorescence microscopy of flake surface showing skin thickness at S and semi-circular pipette imprint. *d*, Epifluorescence microscopy of flake from Station 1 showing dark regions not covered with light coloured skin but speckled with amorphous fluorescent material. *e*, Fortuitous freeze-etched electron micrograph of Station 1 flake skin with embedded diatom tests (D) and hundreds of tiny plates shown at higher magnification in *f*. *g*, Oil immersion epifluorescence microscopy of Station 1 flake surface showing exposed emulsion and a microbial colony (M) in a crevice of the flake. *h*, Freeze-etched electron micrograph of the interior of a Station 1 flake.

Table 1 Physical and chemical characteristics of Ixtoc 1 mousse taken from the surface of the Gulf of Mexico at different times and locations following the 3 June 1979 blowout

Station*	Sample	Distance from wellhead (km)	Maximum time of weathering (days)	Average weight (mg)	Average density (gm cm ⁻³)§	Chemical composition (% total)		
						Aliphatics <i>f</i> ₁	Aromatics <i>f</i> ₂	Polars <i>f</i> ₃
1	Flakes	1,025	79	5.1 (<i>n</i> = 122)	1.017	49.9	16.1	34.0
2	Flakes	875	75	3.5 (<i>n</i> = 44)	1.014	48.6	23.1	28.4
3	Flakes	758	77	1.7 (<i>n</i> = 171)	1.011	44.0	23.2	32.7
					Mean ± s.d.	47.5 ± 3.1	20.8 ± 4.1	31.7 ± 2.9
1	Pancakes	1,025	79	4200 (<i>n</i> = 15)	1.015	56.5	27.7	15.7
2	Pancakes	875	75	87,000 (<i>n</i> = 6)	1.011	58.7	16.5	24.8
					Mean ± s.d.	57.6 ± 1.6	22.1 ± 5.6	20.2 ± 6.4
1	Pancake skin†	1,025	79	<0.1	1.022	50.6	19.0	30.4
4	Wellhead mousse	7	7	—	—	55.7	25.0	19.3
4	Wellhead oil-in-water dispersion‡	<1	<1	—	—	49.3	42.6	8.1

* Station 1: 21 August 1979, 27°29' N, 96°44' W; Station 2: 17 August 1979, 26°10' N, 96°24' W; Station 3: 19 August 1979, 24°10' N, 97°37' W; Station 4: 17 September 1979, 19°24' N, 92°12' W.

† Pancake skin corresponds to the 5–10 μm skin shown in Fig. 1c, d. It was isolated from pancakes by shaking them in seawater. Flakes produced floated with the pancakes while the tiny pieces of skin sank to the bottom, indicating a density greater than seawater (>1.022).

‡ Oil collected at the wellhead with a GO-Flo bottle before it had reached the surface ('reference' wellhead mousse).

§ Density was measured by placing samples in solutions of different salinities. All samples sank at salinities 15‰ (1.009 g cm⁻³).

|| Fractions were isolated by silica gel column chromatography²⁷: Aliphatics—hexane eluate; aromatics—hexane/methylene chloride (50:50) eluate; polars—methanol eluate.

pelagic tar has a single paraffinic maximum in its gas chromatogram which is thought to be typical of weathered crude oils^{16,17}. The distinction between this type of pelagic tar and Ixtoc 1 mousse which also exhibits a single paraffinic maxima (Fig. 2c) is not clear. The presence of a stable water-in-oil emulsion seems to characterize mousse. Mousse pancakes and flakes from Stations 1 and 2 averaged 45–53% water as determined by centri-

fugation (after heating the mousse in a boiling water bath for 10 min). Water contents of pelagic tars have rarely been measured but presumably the waxy tanker precipitates contain little if any water, and the other so-called weathered crudes with single paraffin maxima may have been emulsion or mousse. It is not known whether mousse can become non-emulsified pelagic tar nor how crude oil composition effects emulsion formation

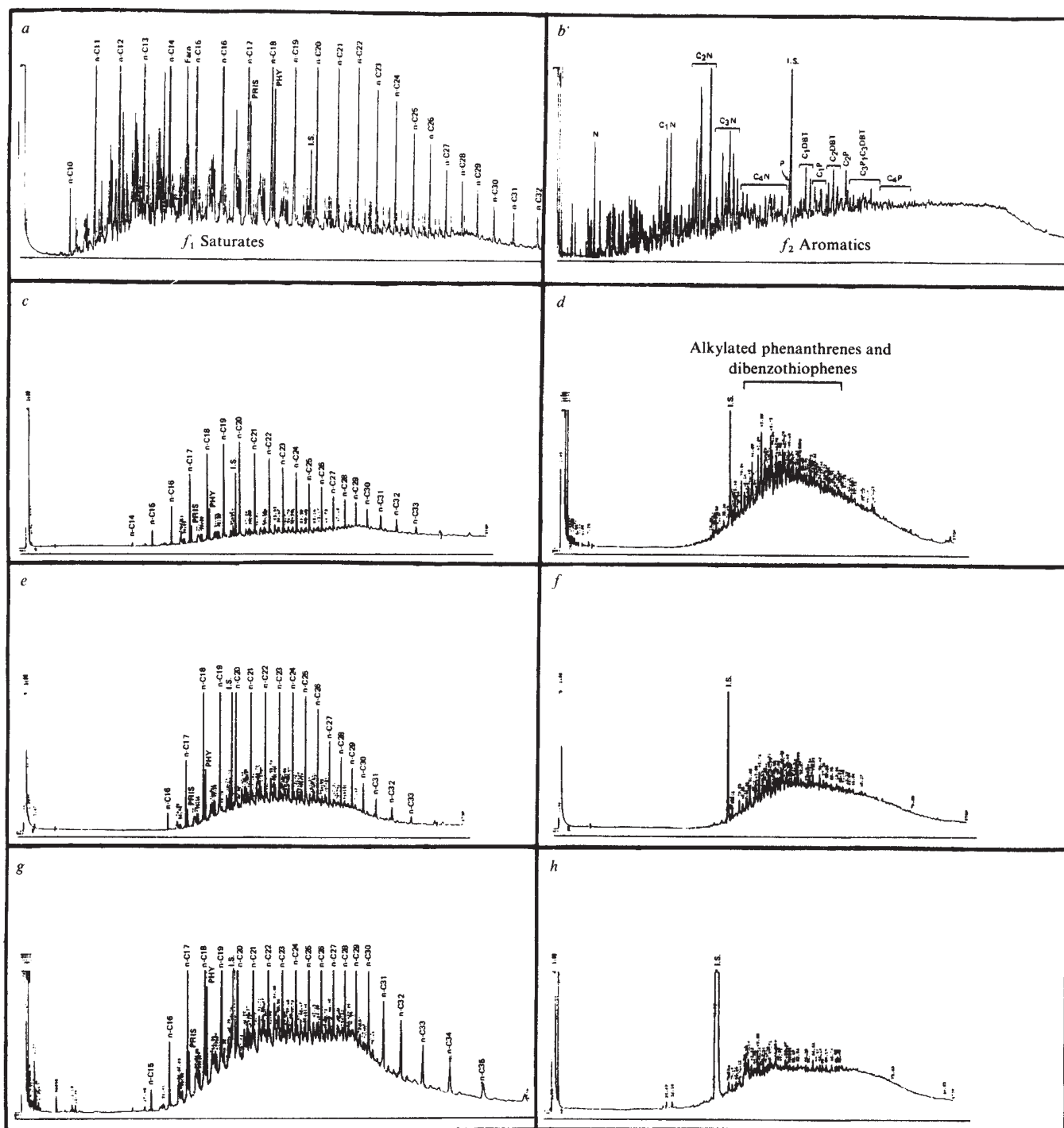


Fig. 2 Glass capillary gas chromatograms of f_1 , saturated (left column) and f_2 aromatic (right column) hydrocarbons from different Ixtoc 1 mousse fractions. An unresolved complex mixture (UCM) appears as a hump below the peaks in most of the chromatograms. *a, b*, Fractions from 'reference' wellhead oil obtained 0.4 km from the blowout (n -C₁₀, n -C₁₁, and so on represent n -alkanes; farn, farnesane; pris, pristane; phy, phytane; N, naphthalene; P, phenanthrene; DBT, dibenzothiophene C₁N, C₂N, C₁P C₁DBT and so on represent alkyl-substituted aromatics; GC, Hewlett Packard 5840A, 30-m SE-20 fused silica (J & W Scientific) splitless injection; temperature, 40–275 °C at 3° min⁻¹. Helium carries gas 1 cm³ min⁻¹ (*c, d*). Gas chromatograms of mousse pancakes obtained at Station 2 along the Texas coast ~875 km from the blowout site (see Table 1). IS, internal standard; (f_1 , androstane; f_2 , deuterated anthracene). *e, f*, Gas chromatograms of mousse flakes from Station 2. *g, h*, Gas chromatograms of Station 1 skin (see Table 1).

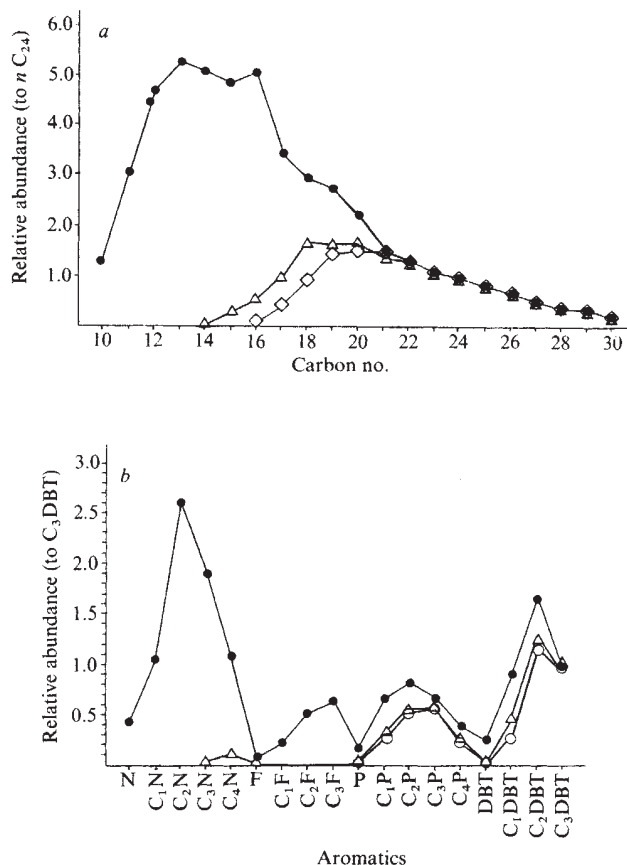


Fig. 3 Relative compositions of wellhead mousse, Texas mousse pancakes and mousse flakes: a, saturates; b, aromatics. ●, Well-head mousse; △, mousse pancakes; ○, mousse flakes.

and stability; some crude oils may even be incapable of forming pelagic tars because of low asphaltene content¹⁶.

Our microscopic examinations suggest that biological factors are not directly involved in causing Ixtoc 1 mousse to crack and flake. Extensive microbial colonization could have been inhibited by the rate of flaking, by photochemical oxidation products^{19,20} or by the lack of nutrient²¹ in the surface waters of the Gulf. Evaporative weathering and physical fragmentation are thought to be the principal processes affecting pelagic tar residues at sea²²⁻²⁴ and wave motion has been postulated as a mechanism for mousse fragmentation^{25,26}. Our observations suggest that weathering processes first form a cracked, scaly surface on pancakes which then 'flake' in turbulent sea. The flaking process exposes fresh mousse within the pancakes and this relatively unweathered mousse creates surface sheen which emanates from the pancake until a skin forms. The skin flakes off on agitation thus ultimately breaking down the mousse pancake as well as altering its composition.

Flaking of stable, water-in-oil emulsion particles seems to be a significant intermediate in the open-sea dispersion of the Ixtoc 1 oil spill. The process causes an increase in the surface-to-volume ratio of spilled oil which would accelerate surface-dependent evaporative dissolution and photochemical weathering and enhance the rate of microbial colonization. The increased density of the flakes relative to pancakes means that their lifetime in the water column may be shortened. The persistence of floating pancakes may be short-lived because of flaking.

We thank John Robinson, John Farrington, Donald Atwood, Pat Parker, Tony Amos and the scientists and crews of the RV Longhorn and RV Researcher for encouragement, constructive criticism and support. This work was supported by NOAA and a small grant from the Georgia Research Foundation.

Received 20 August; accepted 16 December 1980.

1. *Oil Spill Intelligence Reports* Vol. 2, No. 22 (1979)–Vol. 3 No. 13 (1980).
2. Spopper, M. F. *Mar. Pollut. Bull.* **9**, 281–284 (1978).
3. Butler, J. N. *Ocean. Ind.*, October 101–112 (1978).
4. Sturges, W. & Blaha, J. P. *Science* **192**, 367–369 (1976).
5. Molinari, R. L. *J. phys. Oceanogr.* **8**, 779–784 (1978).
6. Pequegnat, L. H. *Contr. mar. Sci.* **22**, 31–39 (1979).
7. Kleinhofs, A., Owais, W. M. & Nilan, R. A. *Mutat. Res.* **55**, 165–195 (1978).
8. Eley, D. D., Key, M. H., Symonds, J. D. & Willison, J. H. M. *J. Coll. Interface Sci.* **54**, 462–467 (1976).
9. Boehm, P. D., Fiest, D. L., Mackay, D. & Paterson, S. in *Proc. 1981 Oil Spill Conf.* (American Petroleum Institute, 1981).
10. Overton, E. B. et al., in *Proc. Conf. on the Preliminary Scientific Findings of the Researcher/Pierce cruise to the Ixtoc 1 Blowout* (NOAA, Washington DC).
11. Horn, M. H., Teal, J. M. & Backus, R. H. *Science* **168**, 245–246 (1970).
12. Heyerdahl, T. *Biol. Conserv.* **3**, 164–167 (1971).
13. Zolsnay, A. *Mar. Pollut. Bull.* **8**, 116–117 (1970).
14. *Petroleum in the Marine Environment*, 48 (National Academy of Sciences, Washington DC, 1975).
15. Eagle, G. A., Green, A. & Williams, J. *Mar. Pollut. Bull.* **10**, 321–325 (1979).
16. Jeffrey, L. M., Peguegnat, W. E., Kennedy, E. A., Vos, A. & James, B. M. *NBS Spec. Publ. Marine Pollution Monitoring* **409**, 233–235 (1974).
17. Butler, J. N. & Harris, J. C. *Mar. Chem.* **3**, 1–7 (1975).
18. Butler, J. N., Morris, B. F. & Sass, J. *Spec. Publ. Bermuda Biological Station for Research* **10**, 1–346 (1973).
19. Larson, R. A., Bott, T. L., Hunt, L. L. & Rogenmuser, K. *Envir. Sci. Technol.* **13**, 965–969 (1979).
20. Hansen, H. P. *Mar. Chem.* **3**, 183–188 (1975).
21. Perry, J. J. *Microbiol. Rev.* **43**, 59–72 (1979).
22. Wade, T. L. & Quinn, J. G. *Mar. Pollut. Bull.* **6**, 54–57 (1975).
23. Butler, J. N. *Mar. Chem.* **3**, 9–21 (1975).
24. Morris, B. F., Butler, J. N., Sleeter, T. D. & Cadwallader, J. *Rapp P.-V. Reun. Cons. Int. Explor. Mer.* **171**, 107–116 (1977).
25. Audunson, T. *Mar. Envir. Res.* **3**, 35–61 (1980).
26. Forrester, W. D. J. *Mar. Res.* **39**, 151–170 (1971).
27. Farrington, J. W. & Tripp, B. W. in *Marine chemistry in the Coastal Environment* (ed. Church, T. M.) 267–284 (American Chemical Society, Washington DC, 1975).

Annual laminations in the sediments of Loe Pool, Cornwall

H. L. K. Simola

Karelian Institute, University of Joensuu, SF-80101 Joensuu 10, Finland

M. A. Coard & P. E. O'Sullivan

Department of Environmental Sciences, Plymouth Polytechnic, Drake Circus, Plymouth, Devon PL4 8AA, UK

Annually laminated freshwater lacustrine sediments have been recorded at several sites in central Europe¹⁻⁵, North America⁶⁻¹³ and Fennoscandia¹⁴⁻²⁴. The presence of laminations may reflect either (1) regular changes within the lake ecosystem itself or (2) variation in the intensity of erosion and transport of material from the catchment, particularly where instability in the lake-watershed system has occurred as a result of human activities²². The principal cause of lamination is, therefore, seasonal variation of environmental conditions, particularly climate. Lakes with laminated sediments tend to be physically deep, exhibit a strong seasonal stratification, and be situated in areas of continental climate. We describe here what we believe to be the first reported instance of a long sequence of laminated lake sediments from Great Britain. Unlike most of the previous examples, these have been formed in a shallow, polymictic lake, in an oceanic climate.

Loe Pool (Fig. 1) is a eutrophic freshwater lagoon at ~4 m OD, 1 km south of Helston in south-west England (GR SW 648250, lat 50°4' N, long 5°17' W) with an area of ~44 hectares and a mean depth of 4 m. Its catchment covers ~50 km², and is mainly farmland, with one major settlement (Helston, population ~10,000). The main stream entering the pool is the River Cober, which drains ~90% of the total catchment. In the nineteenth and early twentieth century this area was the site of extensive mining operations, principally for tin²⁵.

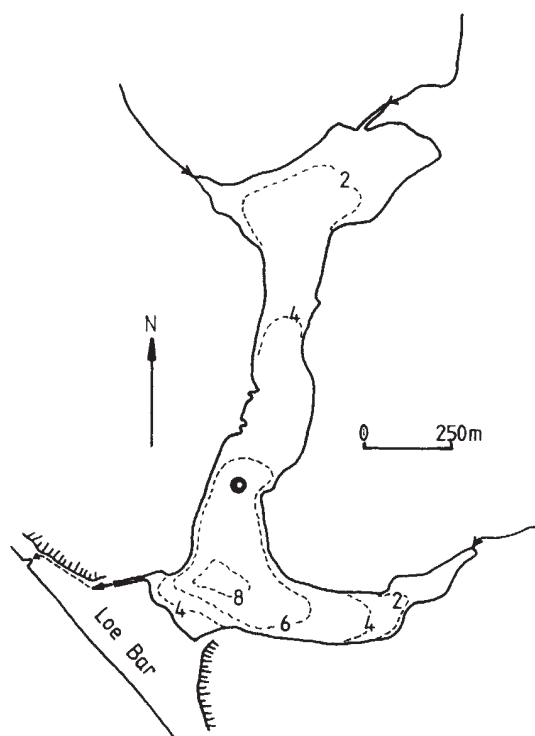


Fig. 1 Loe Pool, Helston, Cornwall, showing depth contours (dashed line) and position of coring site (●).

Three sediment cores were taken from beneath ~7 m of water at the location shown in Fig. 1, where studies (M.A.C. unpublished data) indicate that an undisturbed conformable sequence of sediments is available. One was obtained using a 3-m version of the Mackereth corer²⁶ and two with 1-m *in situ* freezing device²⁷. The Mackereth core was frozen at ~-20 °C to preserve lamination. The surfaces of all cores were then cleaned while still frozen to facilitate description and photography of sediment structures. The stratigraphy of the three cores was easily matched by using prominent marker horizons and characteristic sequences of laminations.

The cores were then allowed to dry in the freezer for a few days to allow adhesive tape preparations to be made¹⁸. These enable the fine structure of the sediment to be examined,

especially for changes in abundance of sub-fossil diatoms. The stratigraphy of the sediment at the sampling site was:

0–40 cm: Highly organic *gyttja*, with four or five pairs of light and dark brown laminations just below the sediment surface (0–5 cm), the rest more homogeneous.

40–120 cm: Irregularly laminated sequence containing red and grey clays alternating with darker layers.

120–300 cm: More regularly laminated sediment consisting of paired grey and black layers, average thickness 3 cm per pair.

The sections 0–7.2 cm, 40–110 cm, 194–206 cm and 230–242 cm were examined. Remains of diatoms, other algae and vivianite crystals were recorded in consecutive 200-μm fields. We shall concentrate here on information obtained from the first and third sections, results being presented in Figs 2 and 3.

In all sections studied, diatom and other algal taxa appear in repeating sequences, the cyclic nature of which we attribute to seasonal algal production and sedimentation. It is therefore possible to define annual increments of deposition, which correspond closely with clearly visible stratigraphic changes. We can therefore identify laminations in these sediments which are truly annual in their nature. Interruptions to the sequence are mainly associated with the deposition of layers of clay, sometimes massive, in which diatoms are relatively scarce, and which we consider to represent dilution of diatom influx by allochthonous, clastic material.

For example, in Fig. 2 (0–7.2 cm), six phases (a–f) are defined. In each of these, *Thalassiosira pseudonana* and *Cyclotella meneghiniana* are succeeded by *Asterionella formosa*, *Melosira granulata* var. *angustissima* and then *Melosira varians*. In monthly plankton sampling of the pool by one of us (M.A.C.) in 1979, the first two taxa were prominent during spring and early summer, and the *Melosira* species in the autumn and winter. It is therefore considered that this sedimentary sequence records seasonal diatom succession in the pool, the *Thalassiosira*/*Cyclotella* stages representing the spring and summer, the *Asterionella*/*Melosira* stages the autumn and winter. The appearance of *Fragilaria* spp. and 'other' (mainly sessile) diatoms in layers thought to represent the winter, would then be consistent with the idea of increased influx of sediment from the littoral and the catchment during these months. Similarly, the summer abundance of *Pediastrum boryanum* in lake plankton samples is recorded in summer laminations. Crystals of vivianite ($\text{Fe}_3(\text{PO}_4)_2 \cdot 8\text{H}_2\text{O}$) seem to be concentrated in some winter layers. This has also been observed in Lovöjärvi (Finland)^{17,18}.

In this section, summer layers coincide with the lighter bands of sediment, and winter layers with the darker. Below 5 cm, however, the sediment structure and the diatom peaks are less

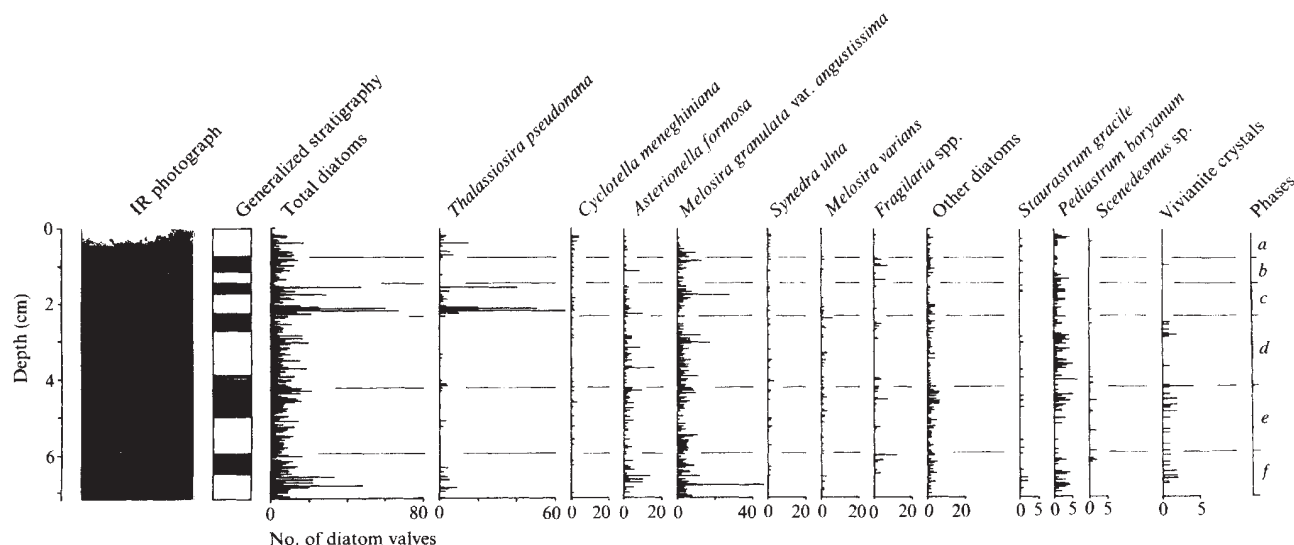


Fig. 2 IR photograph of and results of diatom, other algal and vivianite crystal counts from the section 0–7.2 cm of frozen sediment from Loe Pool.

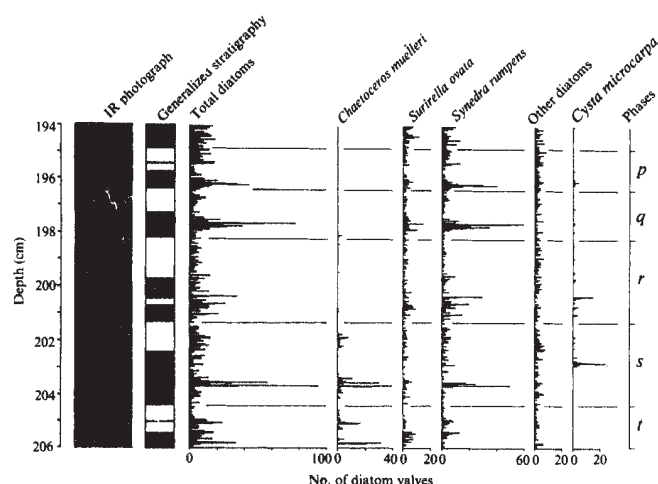


Fig. 3 IR photography, and results of diatom and *Cysta microcarpa* counts from the section 194–206 cm of a frozen sediment core from Loe Pool. Irregular, diagonal striations are caused by ice-crystal formation during freezing of the Mackereth core.

well defined. Inspection of these and other cores indicates considerable bioturbation of the sediments between 5 and 40 cm.

Figure 3 (194–206 cm) shows a section in which five peaks of diatom deposition (*p–t*) are separated by relatively barren layers. The most abundant taxa present are *Chaetoceros muelleri*, *Surirella ovata* and *Synedra rumpens*, which tend to appear in a sequence which denotes seasonal succession. On three occasions, maxima of these planktonic diatoms are immediately followed by peaks of *Cysta microcarpa* (*sensu* Nygaard)²⁸. We interpret the planktonic diatom peaks, which coincide with darker bands of sediment, as representing the summer months, and the relatively barren layers, which correspond to the grey-brown clays, and in which 'other' diatoms largely occur, as the winter.

Figure 4 clearly shows that despite the oceanicity of the climate of this part of Europe, a pronounced winter maximum of rainfall and stream flow occurs, and that summer temperatures substantially exceed those of the winter. This contrast between the seasons seems sufficient to account for the formation of the Loe Pool laminations.

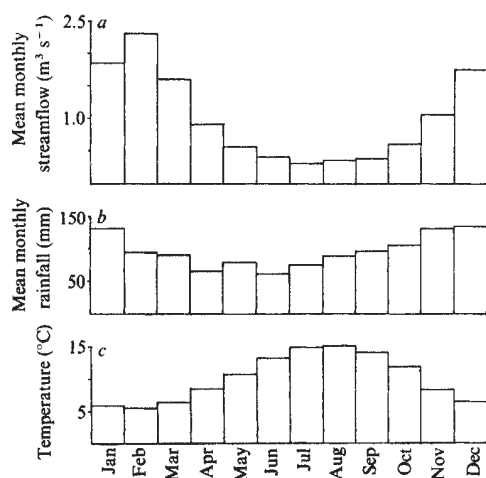


Fig. 4 *a*, Mean monthly streamflow (River Cober, 1970–79); *b*, mean monthly rainfall (1941–70); and *c*, mean monthly temperature (1960–74) for the Loe Pool catchment. Streamflow data from the South West Water Authority gauging station at Helston, rainfall data from Wendron, and temperature data from RNAS Culdrose.

From 300 to ~120 cm these consist of pairs of black (summer) and grey-brown (winter) layers, recording increased inflow of clastic material in the winter months, covering a period of ~60 yr. Then, between 120 and 40 cm, massive clay layers occur, which, according to diatom analysis not described in detail here (H.L.K.S. and M.A.C., in preparation), cover a period of only 7 yr. Here a high proportion of the sediment consists of material originating as mining wastes, and peaks in planktonic diatoms are separated by as much as 20 cm of clay in which diatom influx has been considerably diluted, thus indicating very rapid accumulation. We correlate this sequence with the most recent period of active mining in the catchment which ended in AD 1938. Within this section, two layers of red, haematite-rich clay occur, the upper at 40–46 cm, the lower at 65–70 cm. Documentation²⁹ shows that during the 1920s and 1930s the pool was heavily polluted by mining wastes, often to the extent that discoloration of the waters occurred. We therefore conclude that the red clay layers were deposited during this time. Between 1860 and 1920 active, but less intensive mining

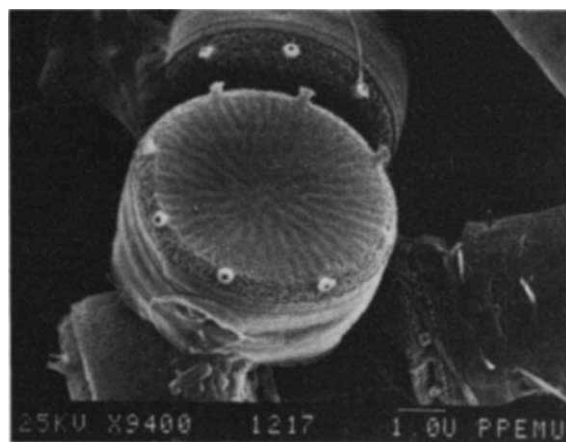


Fig. 5 Scanning electron micrograph of *Thalassiosira pseudonana* Hasle et Heimdal from Loe Pool plankton samples. (Photograph Martin Coard.)

took place²⁵. Above 40 cm, laminations are not observed until ~5 cm from the sediment surface. Here there are four or five laminations which are not visually prominent, but which are well defined in the algal stratigraphy. Between 5 and 40 cm the sediment is considerably bioturbed, but ¹³⁷Cs analysis shows that the 1963 peak lies between 14 and 16 cm depth. This suggests a mean accumulation rate for the section between 16 cm and the sediment surface of 1 cm yr⁻¹, which is in agreement with the rate calculated from lamination counts just below the sediment surface.

The section 0–40 cm, between the top of the upper haematite clay and the sediment surface, thus represents the period since 1938, and has a mean accumulation rate very close to 1 cm yr⁻¹. Together with a sedimentation rate of 80 cm in 7 yr for the section 40–120 cm (based on diatom stratigraphy), and of 3 cm yr⁻¹ for the section 120–300 cm (based on lamination counts), this estimate allows us to conclude that the Loe Pool cores cover a period of ~110 yr.

Between ~1870 and ~1920 the main cause of lamination seems to have been a steady inflow of mining wastes in winter. During the period of most intensive mining (1920–38) this increased to massive proportions (80 cm in 7 yr). Between 1940 and 1975 laminations were not formed. The presence of the remains of numerous burrows indicates that this was a phase of considerable benthic activity. The formation of laminations in very recent years is attributed to increased eutrophy of the pool, which has led to occasional instances of anoxia at the sediment

surface during the summer months. These most recent laminations are therefore being produced in different conditions from those formed at an earlier date, which may in part account for colour difference between the respective summer and winter layers.

We thank Lieutenant Commander J. P. Rodgers and the National Trust for permission to carry out work at Loe Pool. We also thank R. S. Cambray for ^{137}Cs analysis and Dr A. C. Hamilton for the loan of a Mackereth corer. The work was carried out while H.L.K.S. and M.A.C. were supported by Devon Local Authority Research Assistantships, and H.L.K.S. by a grant from the Alfred Kordelin Foundation (Finland).

Received 10 October 1980; accepted 14 January 1981.

1. Nipkow, F. Z. *Hydrol. (Budapest)* **1**, 1–23 (1920); *Schweiz. Z. Hydrol.* **4**, 71–80 (1927).
2. Thomas, E. A. *Vjschr. naturf. Ges. Zürich* **109**, 103–142, (1964).
3. Geyh, M., Merkt, J. & Müller, H. *Arch. Hydrobiol.* **69**, 366–399 (1971).
4. Thompson, R. & Kelts, K. *Sedimentology* **21**, 577–596 (1974).
5. Kelts, K. & Hsü, K. M. in *Lakes: Geology, Chemistry, Physics* (ed. Lerman, A.) (Springer, Berlin, 1979).
6. Tippet, R. *Can. J. Bot.* **42**, 1693–1709 (1964).
7. Ludlam, S. D. *Limnol. Oceanogr.* **14**, 848–857 (1969); **21**, 743–746 (1976).
8. Craig, A. J. *Ecology* **53**, 46–57 (1972).
9. Anthony, R. S. *Limnol. Oceanogr.* **22**, 45–54 (1977).
10. Swain, A. *Quat. Res.* **3**, 383–396 (1973); **10**, 55–68 (1978).
11. Edmondson, W. T. *Verh. int. Ver. Limnol.* **19**, 770–775 (1975).
12. Allison, D. E. *Limnol. Oceanogr.* **15**, 138–144 (1970).
13. Cwynar, L. C. *Can. J. Bot.* **56**, 10–21 (1978).
14. Digerfeldt, G., Battarbee, R. W. & Bengtsson, L. *Geol. För. Stockh. Förh.* **97**, 29–40 (1975).
15. Renberg, I. *Geol. För. Stockh. Förh.* **98**, 335–360 (1976); *Early Norrland* **11**, 63–92 (1972).
16. Kukkonen, E. & Tynni, R. *Acta bot. fenn.* **90**, 1–30 (1971).
17. Saarnisto, M., Huttunen, P. & Tolonen, K. *Ann. Bot. Fenn.* **14**, 35–45 (1977).
18. Simola, H. *Ann. Bot. Fenn.* **14**, 143–148 (1977).
19. Vuorinen, J. *Pol. Arch. Hydrobiol.* **25**, 443–451 (1978).
20. Tolonen, M. *Ann. Bot. Fenn.* **15**, 177–208, 209–222, 223–240 (1978).
21. Koivisto, E. & Saarnisto, M. *Geogr. Ann.* **A60**, 55–61 (1978).
22. Simola, H. *Holarctic Ecol.* **2**, 160–168 (1979).
23. Tolonen, K. *Boreas* **9**, 11–19 (1980).
24. Battarbee, R. W., Cronberg, G. & Lowry, S. *Hydrobiologia* **71**, 225–232 (1980).
25. Hamilton-Jenkins, A. K. *Wendron Tin* (Wendron Forge Ltd, Helston, 1978).
26. Mackereth, F. J. H. *Limnol. Oceanogr.* **3**, 181–191 (1958).
27. Huttunen, P. & Meriläinen, J. *Ann. Bot. Fenn.* **15**, 128–130 (1978).
28. Berg, K. & Petersen, I. *Folia limnol. scand.* **8**, 1–273 (1956).
29. *Western Independent* Penzance (2 February 1928).

Supply of iron and manganese to an anoxic lake basin

W. Davison

Freshwater Biological Association, Windermere Laboratory, Ambleside, Cumbria LA22 0LP, UK

In seasonally anoxic lakes, the dynamic cycling of iron and manganese makes possible the study of geochemical cycling and processes of mineral formation, while the rapid lake processes may be a model for the analogous but slower oceanic processes. We have now measured concentration–depth profiles for iron and manganese to assess the flux of these elements from the sediments. The seasonal accumulation of iron in lake water can be accounted for by supply from the sediment, but the manganese profiles suggest that the major source of manganese is dissolution from particular material sedimenting through the water column.

This research is part of a concerted programme^{1,2} on the geochemical cycling of iron and manganese in lakes. The characteristic features of the process in lakes subject to deep-water anoxia are well known^{3–5}. Although the observed distribution of the various forms of iron in such conditions has recently been related to individual rate processes on the basis of detailed measurements¹, no such treatment exists for manganese. However, the concentration–depth profiles usually have a significantly different shape from those of iron, and this has been attributed³ to manganese being more easily reduced. It has been assumed^{3,6,7} that manganese is supplied from the sediment in anoxic conditions. Golterman⁸ speculated that

manganese oxides may be reduced in the water column to Mn(II) with concomitant oxidation of Fe(II), and Delfino and Lee⁹ realized that dissolution or desorption processes in the water column could be significant to the generation of Mn(II). Despite detailed studies of manganese in seasonally anoxic lakes^{9–11} the roles of dissolution of particulate matter or release from the sediment have not been clarified. However, most workers have looked at ‘macro-profiles’—concentrations measured at depth intervals of 1 m or greater—whereas to study sediment–water interactions ‘micro-profiles’—concentrations measured at centimetre distances from the sediment—need to be known.

Water samples were collected from the seasonally anoxic Esthwaite Water^{1,4,5} at approximately monthly intervals during the summer of 1980. A pump incorporating a sediment detector¹² was used to collect samples for both micro- and macro-profiles. The samples were analysed for Fe(II) and Mn(II) by polarography¹³, and for total iron and manganese by acid digestion procedures^{14,15}.

Figure 1 shows the results with respect to distance from the sediment and time of the year. Regression analysis of the micro-profiles for both total and soluble manganese measured between 5 and 50 cm from the sediment showed that the change in concentration with respect to distance from the sediment (dc/dz) did not significantly differ from zero ($P > 0.8$). The mean of the 11 values for dc/dz was $-0.87 \mu\text{g l}^{-1} \text{cm}^{-1}$ (range -1.96 to $+0.52$). The slopes calculated from the macro-profiles, 0.5–1.5 m from the bottom showed that the total and polarographic measurements gave similar results and the value of dc/dz was $\sim -0.5 \mu\text{g l}^{-1} \text{cm}^{-1}$.

In contrast to manganese there is no polarographically measurable Fe(II) during May and June and it is only just measurable in early July (Fig. 1). During this time the total iron in the water column increases substantially. This has been explained^{1,4,5} by assuming that the oxidized sediment layer, which during the winter months overlies the very reducing sediment, is itself reduced when the lake stratifies and so Fe(II) diffuses into the water column. The overlying water is not then entirely reducing and so Fe(II) is oxidized almost immediately, and the resultant Fe(III) therefore builds up in the water column. When Fe(II) does appear in the water column in July and August a greater concentration gradient exists in the 25 cm of water nearest the sediment. Using regression analysis to calculate the slope reveals that the flux calculated from the micro-profiles was 10–30 times greater than that estimated from the slope of the macro-profile at 0.5–1.5 m.

Davison *et al.*¹ have shown that the lateral flux of iron supplied from sediment at the same depth as the water must be taken into account. Moreover, towards the latter part of the season, dissolution of particulate iron falling through the water column also becomes important. Thus the macro-profile cannot be used for estimating the direct flux from the sediment. However, from Fig. 1 it seems that in the confined zone of the bottom 25 cm of water the chemical gradient from the sediment dominates the lateral component of water movement and so a gradient which reflects the sediment–water interaction is established.

The seasonal increase in the iron and manganese in the bottom waters of the lake varies little from year to year^{1,4,5}. For 1979, 1.73 and 1.34 tonnes of iron and manganese respectively had accumulated in the larger lake basin by the end of the season. Assuming that both elements may dissolve for a period of 12 weeks from all that sediment which may possibly become anoxic, that is below 8.5 m depth, the mean micro-profile value for 23 July 1980 and 14 August 1980 of $-82 \mu\text{g l}^{-1} \text{cm}^{-1}$ for iron, results in 2.20 tonnes supplied from the sediment. For manganese, however, a slope of $-1 \mu\text{g l}^{-1} \text{cm}^{-2}$ results in 0.027 tonnes, only 2% of the manganese found to have accumulated in the lake. Therefore although iron may be sediment derived, even if a previously measured manganese concentration gradient, which is 5 times the value used here, is considered¹², manganese is not. This is verified by the shape of the manganese profiles. The virtual invariance of concentration with respect to depth in the stratified bottom waters indicates a lack of any

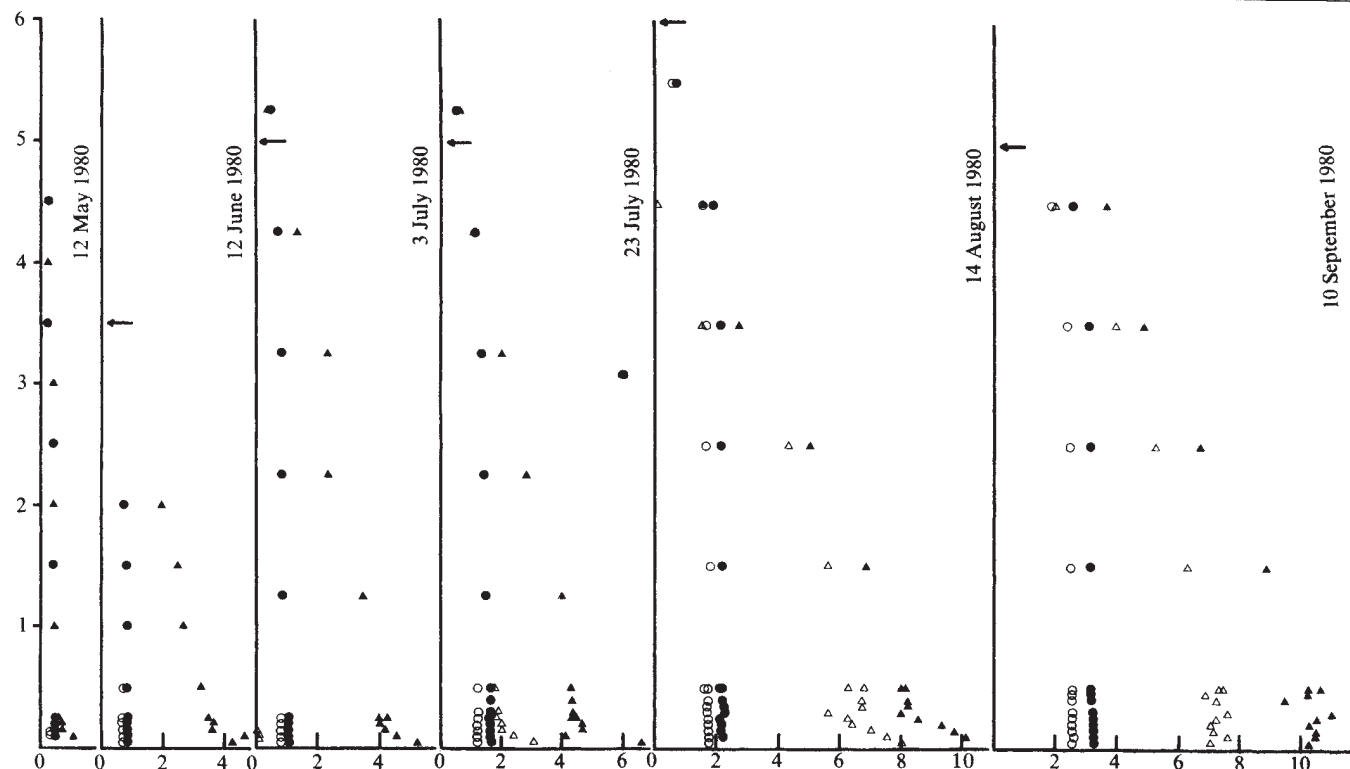


Fig. 1 The concentration of polarographically measured iron (Δ), total digested iron ($\blacktriangle$), polarographically measured manganese ($\circ$) and total manganese ($\bullet$) versus distance from the sediment. The position of the oxycline below which oxygen is not measurable ($<0.3 \text{ mg l}^{-1}$) by an *in situ* probe¹⁶ is indicated. On 12 May 1980 oxygen was still present at 4.8 mg l^{-1} near the sediment.

active point source or sink. Yet the progressive increase in concentration with time indicates a uniform supply of manganese at all depths. Comparison with the iron profiles shows that the shape cannot be accounted for by supply from lateral sediments; no similar increase in the waters in close proximity to the sediment is evident. On the macro-scale, supply from lateral sediment could not produce such a uniform profile because the ratio of sediment area-to-volume at any given stratum increases with depth.

Profile shapes show that at most 50% of the flux of iron at any depth is due to dissolution of particulate material¹. The calculation¹⁷ uses the concentration gradient and a knowledge of the vertical eddy diffusion coefficient to estimate the flux between each horizontal stratum of water which was assumed to be uniformly mixed. The lateral sediment derived flux was taken into consideration and any flux imbalance was attributed to *in situ* dissolution processes. Using the profiles of 14 August 1980 and 10 September 1980 for manganese the calculation suggests that dissolution accounts for 80–90% of the manganese flux. Although the approximations involved mean the results are semi-quantitative, the trend offers a possible explanation for the shape of the manganese profile. Particulate manganese from the inflowing streams must always be falling through the water column and may act as a source of supply for soluble manganese at all depths. The estimated 1.3 tonnes of manganese which accumulates in the hypolimnion results in a large increase in concentration in these bottom waters, but its contribution to the total manganese in the lake basin is not so great because of the relatively large volume of water in the surface layers. Before stratification uniform manganese concentration profiles with $100 \mu\text{g l}^{-1}$ of Mn could be measured, suggesting that 0.4 tonnes of manganese were present in the lake before any accumulation began. Moreover, it was estimated by gauging that $6 \times 10^6 \text{ m}^3$ of water flowed into the lake during June, July and August of 1980. The scant measurements indicate that the concentrations in the inflows are two to three times higher than in the epilimnion and so this volume of water probably carried ~ 1.2 tonnes of manganese into the lake. Therefore sufficient manganese must

either have been present in the lake or supplied to it, to account for the seasonal hypolimnetic increase.

The rapid increase in both polarographic and total manganese below the oxycline (Fig. 1) shows that the rate of vertical mixing in the metalimnion is not sufficiently rapid to mask the oxidation of the reduced form. Although the profiles of particulate manganese can be deduced from the difference between the two components plotted in Fig. 1, no conclusions can be drawn because the fraction of the particulate concentration which dissolves at any depth will depend on the rate of supply of manganese to the lake, the rate of dissolution, the sinking rate and the gradient of redox intensity. Simple solubility calculations show that neither MnCO_3 nor MnS can control the profile shape^{18,19}.

The build up of manganese in the hypolimnion therefore cannot be simply attributable to dissolution from the sediment. The shapes of the iron and manganese profiles show that simple vertical transport processes cannot account for the near-vertical manganese profiles and *in situ* supply by dissolution of sedimenting particulate material accounts for the observations.

The present results agree well with those of related studies. Eaton²⁰ showed that a considerable flux of the manganese in a fjord resulted from the dissolution of manganese oxides in anoxic conditions. There is evidence from estuarine work^{21–23} for manganese redissolving in the upper estuary where low pH and Eh prevail. The manganese profile in the permanently anoxic Black Sea has been modelled on the supposition of *in situ* dissolution²⁴. The same explanation may apply to a comparable peaked profile reported for a lake²⁵; slight peaks have also been reported for other lakes^{3,26}.

The interpretation of the iron and manganese ratios in the sediments of Esthwaite Water¹¹ has been based on the supposition that a substantial manganese fraction is lost from the sediment during anoxia: these results may require fresh interpretation. Care must be taken in extrapolating the results from artificially enclosed sediment–water systems to lakes^{4,5,27,28} because the possible input of manganese from dissolution of sedimenting particles is excluded. Mortimer^{4,5} reported the

presence of soluble manganese 40 days after Fe(II) had appeared in his laboratory tank experiments, which contrasts with the natural system where soluble manganese appears first.

Interstitial water profiles in sediments have been commonly used to estimate fluxes to the water column²⁸. The rate of transport is usually $\sim 10^3$ times slower than in the hypolimnion studied here. As micro-profiles discriminating to <25 cm are necessary for the hypolimnetic waters, by analogy the resolution of interstitial water measurements needs to be to <0.25 mm because lateral processes such as interchange with labile mineral fractions can still occur. Therefore caution should be exercised in using interstitial water profiles to estimate fluxes through the sediment-water interface.

I thank E. Tipping for constructive criticism and C. Woof for technical assistance.

Received 13 October 1980; accepted 14 January 1981.

1. Davison, W., Heney, S. I., Talling, J. F. Rigg, E. Schweiz Z. Hydrol. (in the press).
2. Tipping, E. *Geochim. cosmochim. Acta* **45**, 191-199 (1981).
3. Hutchinson, G. E. *A Treatise on Limnology* Vol. 1 (Wiley, New York, 1957).
4. Mortimer, C. H. *J. Ecol.* **29**, 280-329 (1941).
5. Mortimer, C. H. *J. Ecol.* **30**, 147-201 (1942).
6. Cole, G. A. *Textbook of Limnology* (Mosby, St Louis, 1975).
7. Wetzel, R. G. *Limnology* (Saunders, London, 1975).
8. Golterman, H. L. *Physiological Limnology* (Elsevier, Amsterdam, 1975).
9. Delfino, J. J. & Lee, G. F. *Envir. Sci. Technol.* **2**, 1094-1100 (1968).
10. Howard, H. H. & Chisholm, S. W. *Am. midl. Nat.* **93**, 188-197 (1975).
11. Mackereth, F. J. H. *Phil. Trans. R. Soc. B* **250**, 165-213 (1966).
12. Cunningham, C. R. & Davison, W. *Freshwat. Biol.* **10**, 413-418 (1980).
13. Davison, W. *Limnol. Oceanogr.* **22**, 746-753 (1977).
14. Davison, W. & Rigg, E. *Analyst* **101**, 634-638 (1976).
15. Mackereth, F. J. H., Heron, J. & Talling, J. F. *Scient. Publ. Freshwat. Biol. Ass.* **36**, (1978).
16. Mackereth, F. J. H. *J. scient. Instrum.* **41**, 38-41 (1964).
17. Hesslein, R. H. *Can. J. Fish. aquat. Sci.* **37**, 552-558 (1980).
18. Holdren, G. R. & Bricker, O. P. *ERDA Symp. Ser.* **42**, 178-196 (1977).
19. Morgan, J. J. in *Principles and Applications in Water Chemistry* (eds Faust, S. D. & Hunter, J. V.) 561-624 (Wiley, New York, 1967).
20. Eaton, A. *Geochim. cosmochim. Acta* **43**, 429-432 (1979).
21. Knox, S. & Turner, D. R. *Estuar. coast mar. Sci.* **10**, 317-324 (1980).
22. Duinker, J. C., Wollast, R. & Billen, G. *Estuar. coast mar. Sci.* **9**, 727-738 (1979).
23. Wollast, R., Billen, G. & Duinker, J. C. *Estuar. coast mar. Sci.* **9**, 161-169 (1979).
24. Spencer, D. W. & Brewer, P. G. *J. geophys. Res.* **76**, 5877-5892 (1971).
25. Ruttner, F. *Arch. Hydrobiol. Suppl.* **8**, 197-454 (1931).
26. Verdouw, H. & Dekkers, E. M. J. *Arch. Hydrobiol.* **89**, 509-532 (1980).
27. Dyrssen, D. & Hallberg, R. *Chem. Geol.* **24**, 151-159 (1979).
28. Lu, J. C. S. & Chen, K. Y. in *Chemistry of Marine Sediments* (ed. Yen, T. F.) 181-221 (Ann Arbor, Michigan, 1977).
29. Murray, J. W. & Gill, G. *Geochim. cosmochim. Acta* **42**, 9-19 (1978).

Methylmercury production in the marine water column

G. Topping & I. M. Davies

Department of Agriculture and Fisheries for Scotland, Marine Laboratory, Victoria Road, Aberdeen AB9 8DB, UK

Although the biosynthesis of methylmercury in sediments is well established¹, this is not necessarily the exclusive natural source of methylmercury entering the marine food chain, particularly commercial fish and shellfish species for human consumption. An examination of mercury levels in freshwater fish², collected from a lake with a history of industrial mercury contamination, suggested that levels in fish are controlled in part by mercury in suspension and it followed that methylation should occur in the water column. Although methylmercury is present in seawater in coastal areas receiving discharges of waste containing either inorganic mercury³ or methylmercury⁴ there is no evidence that methylmercury is actually formed in the water column. We now present data which demonstrate that inorganic mercury can be methylated in the water column and we compare this production with that known to occur in marine sediments.

In May 1979 experiments to investigate the movement of inorganic mercury between the water column and sediments were carried out using a large plastic bag (3 m diameter $\times$ 20 m deep) moored in an arm of a sea loch (Loch Ewe) on the west

coast of Scotland. This type of enclosure, which is closed at the bottom, has been used⁵ to study the effects of low levels ($\sim 1 \mu\text{g dm}^{-3}$) of pollutants, particularly metals, on natural populations of phytoplankton, zooplankton and other trophic levels captured in the bags.

In our study, the seawater in the enclosure was initially spiked with inorganic mercury as HgCl_2 , to a concentration of $\sim 0.5 \mu\text{g dm}^{-3}$ using a method described elsewhere⁶. Seawater samples were analysed for total and 'reactive' mercury⁷. Sub-samples of the total settlement material, which was pumped daily from the bottom of the bag, were analysed for total mercury⁸, methylmercury⁹ and carbon (using a CHN analyser). Moderately active primary production was maintained by regular additions of nutrients to the plastic enclosure giving a

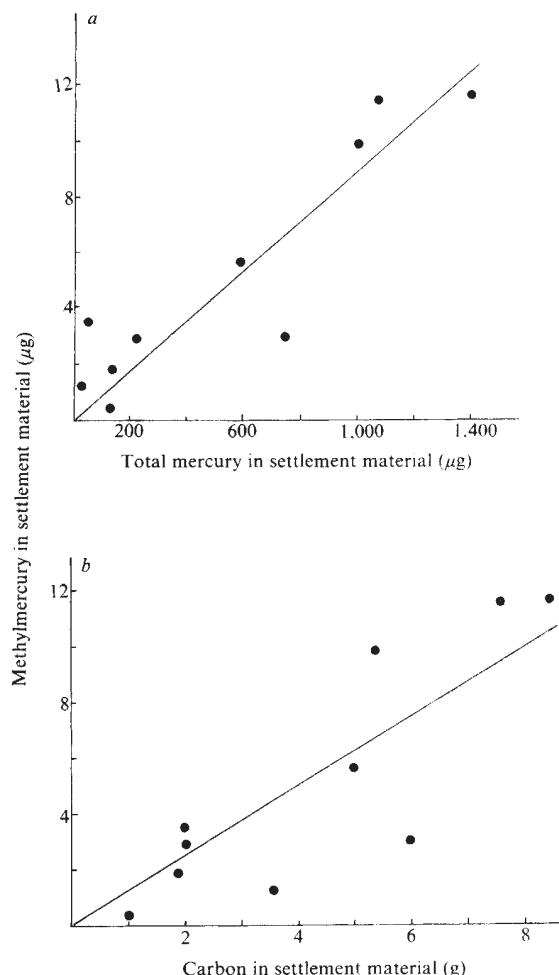


Fig. 1 a, Relationship between methylmercury and total mercury in settlement material. (Regression analysis: $y = 0.00885x$; $r = 0.91$ and standard error of slope = 0.00126.) b, Relationship between methylmercury and carbon in settlement material. (Regression analysis: $y = 1.233x$; $r = 0.85$ and standard error of slope = 0.294.) Both correlations are significant at the 0.01% level.

mean chlorophyll level of 10.9 mg m^{-3} and a mean fallout of organic carbon of $0.4 \text{ g C m}^{-2} \text{ day}^{-1}$.

Significant quantities of methylmercury ($1-10 \mu\text{g day}^{-1}$) were found in the settlement material. These concentrations were strongly correlated with concentrations of organic carbon ($r = 0.85$) and total mercury ($r = 0.91$) in settlement material; both correlations are significant at the 0.01% level (Fig. 1). Because the seawater in the bag was isolated from the natural sediments in the sea loch throughout the experiment and because the settlement material was not allowed to accumulate for a long

period (about 24 h), we believe that a large part of the methylmercury was formed within the water column.

The amount of methylmercury formed in suspension and collected with the settlement material is directly related to the total mercury associated with the particulate organic matter which in turn is dependant on the level of primary production. From Fig. 1 the ratio methylmercury (μg)/carbon (g) is 1.23 and the ratio methylmercury (μg)/total mercury (μg) is 0.0088; the latter may be compared with 0.001–0.003 estimated by Davies *et al.*³ for a contaminated estuary.

The present experimental results can be used to obtain a rough estimate of the importance of the formation of methylmercury in the water column. If the ratio of methylmercury to mercury of 0.0088 is a guide to what happens in the oceans, and assuming a mean concentration of mercury in the oceans¹⁰ of 10 ng dm^{-3} , the total quantity of methylmercury in open ocean water is 121,000 tonnes. From the global production of carbon¹, our measured ratio of methylmercury (μg) to carbon (g), and assuming that the rate of production of methylmercury is proportional to the total mercury concentration in open ocean, we estimate the amount of methylmercury formed each year in association with carbon fixation at 492 tonnes.

The estimation of methylmercury concentration and production in marine sediments is complicated by the wide range of estimates^{11–16} of methylmercury production rates, from 0.1–240 $\mu\text{g cm}^{-2} \text{ yr}^{-1}$. Of the two estimates^{14,15} referring to sediments having low mercury concentrations, one¹⁵ includes both water column and sediment production and the other¹⁴ gives a minimum rate of production of 0.56 $\mu\text{g cm}^{-2} \text{ yr}^{-1}$ based on uptake by plants and animals in the sediments (containing 2 $\mu\text{g cm}^{-3}$ of total mercury). We use the latter value here. We take the mean concentration of mercury in marine sediments¹ as 0.3 $\mu\text{g per g wet weight}$ (0.45 $\mu\text{g cm}^{-3}$) and the fraction of mercury present as methylmercury in sediments¹⁶ as 0.1%. If the relevant volume of the sediment compartment is $3.61 \times 10^{19} \text{ cm}^3$ (corresponding roughly to a 10-cm depth over a sea bottom surface of $361 \times 10^{12} \text{ m}^2$) the total methylmercury in marine sediments is 16,200 tonnes. Similarly, production of methylmercury by sediments amounts to 450 tonnes yr^{-1} .

Knauer and Martin¹⁷ found concentrations of 1.3–4.5 ng per g of methylmercury in selected phytoplankton samples from Monterey Bay, California. Using these data (mean value of 3 ng per g) and an annual primary production of $2 \times 10^{17} \text{ g wet weight}$ we estimate that the total amount of methylmercury associated with the annual global production of phytoplankton is ~600 tonnes. This value is comparable with one we obtained for phytoplankton using data from our bag investigations. By comparison, the quantity of methylmercury (~48 tonnes) associated with the annual global production of fish¹ is small.

These data indicate that not only is the 'standing stock' of methylmercury in the oceans greater than that in the surface sediments but that the annual production of methylmercury in these two compartments is similar. Investigations in the field are now necessary to verify our experimental findings and speculations.

Received 15 October 1980; accepted 7 January 1981.

1. An Assessment of Mercury in the Environment (National Research Council, National Academy of Sciences, Washington DC, 1978).
2. Armstrong, F. A. J. & Scott, D. P. *J. Fish. Res. Board Can.* **36**, 670 (1979).
3. Davies, I. M., Graham, W. C. & Pirie, J. M. *Mar. Chem.* **7**, 111 (1979).
4. Egawa, H. & Tajima, S. in *Proc. 2nd US Japan Experts Meet.*, Tokyo (1976).
5. *Bull. mar. Sci.* **27**, No 1 (1977).
6. Topping, G. & Windom, H. L. *Bull. mar. Sci.* **27**, 135 (1977).
7. Topping, G. & Pirie, J. M. *Analyt. chim. Acta* **62**, 200 (1972).
8. Topping, G. *Proc. Anal. Div. chem. Soc.* **14**, 222 (1977).
9. Davies, I. M. *Analyt. chim. Acta* **102**, 189 (1978).
10. Baker, C. W. *Nature* **270**, 230 (1977).
11. Langley, D. G. *J. Wat. Pollut. Cont. Fedn* **45**, 44 (1973).
12. Olsen, B. H. & Cooper, R. C. *Wat. Res.* **10**, 113 (1976).
13. Shin, E. & Krenkel, P. A. *J. Wat. Pollut. Cont. Fedn* **48**, 473 (1976).
14. Windom, H. L., Gardner, W., Stephens, J. & Taylor, F. *Estuarine Coastal mar. Sci.* **4**, 579 (1976).
15. Davies, I. M., Graham, W. C. & Pirie, J. M. *ICES CM* (Fisheries Improvement Committee, E: 31, 1978).
16. Bartlett, P. D., Craig, P. J. & Morton, S. F. *Sci. Total Envir.* **10**, 245 (1978).
17. Knauer, G. A. & Martin, J. H. *Limnol. Oceanogr.* **17**, 868 (1972).

Chondro-osseous morphology of *Dermochelys coriacea*, a marine reptile with mammalian skeletal features

Anders G. J. Rhodin*†, John A. Ogden* & Gerald J. Conlogue*

* Department of Surgery (Orthopaedics), Skeletal Growth and Development Study Unit, Marine Mammal Stranding and Study Center and the Peabody Museum, Yale University, New Haven, Connecticut 06510, USA

† Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA

The marine leatherback turtle, *Dermochelys coriacea* (Testudines, Dermochelyidae), is the largest known extant reptile, reaching weights >680 kg. It has a cosmopolitan oceanic distribution, nesting in tropical climates but migrating widely into colder temperate and boreal waters¹. Recent studies have shown that the leatherback has certain well developed thermoregulatory adaptations: vascular counter-current heat exchangers in the flippers, thick subcutaneous insulation beneath its leathery shell and effective inertial homoiothermy, if not endothermy^{2,3}. Whether this ability to maintain a body temperature higher than the ambient is due to endogenous endothermy or thermal inertia remains unclear^{4,5}. The skeleton of *Dermochelys* remains extensively cartilaginous even in adult animals^{6,7}, which has been attributed to neoteny secondary to its highly pelagic habitat, based on the assumption that its skeleton resembled that of typical embryonic turtles⁸. However, up to now there have been no studies of the internal architecture of the leatherback appendicular skeleton. We report here that the chondro-osseous morphology of *Dermochelys* is unlike that of any other known extant turtle or reptile but is more similar to that of marine mammals, notably Cetacea (whales) and Sirenia (manatees).

The typical hard-shelled marine turtles (Cheloniidae—genera *Chelonia*, *Caretta*, *Eretmochelys* and *Lepidochelys*) show a pattern of chondro-osseous morphology highly reminiscent of terrestrial or aquatic turtles^{9,10}. A thick, distinct compacta is clearly delineated from a medullary cancellous region and the epiphyses are covered by a very thin avascular cartilage which serves the purpose both of articular cartilage and physis, or growth plate. The subchondral plates of the epiphyseal bone surfaces are smooth and perfectly parallel to the articular cartilages. The only difference in marine turtles is the failure of development of a medullary cavity.

Dermochelys shows several important differences in skeletal morphology. In dried whole bones (with cartilage removed) the epiphyseal subchondral plates are not smooth, but instead present a rough, undulated, fenestrated surface which is not parallel to the contours of the articular cartilage surface (Fig. 1). In longitudinal sections of fixed whole bones each epiphyseal end has large cartilaginous epiphyses. Each of these, whether large or small, is filled with an extensive cartilage canal vascular system (Fig. 1). The vessels are both perichondral and transphyseal in origin and allow continuity of circulation between metaphysis and epiphysis. Both large and small vessels cross the physis: the smaller ones probably participate in endochondral chondro-osseous replacement, and the large ones are probably involved in cartilaginous expansion and nutrition of the epiphysis itself. No secondary calcification or ossification centres develop in any of these cartilaginous areas.

The diaphysis and metaphysis are filled with relatively dense cancellous trabecular bone without secondary medullary cavity formation. The metaphysis seems to be almost totally derived from endochondral bone formation and the diaphysis from a combination of endochondral and periosteal membranous growth. This is dramatically evident because of a combination of

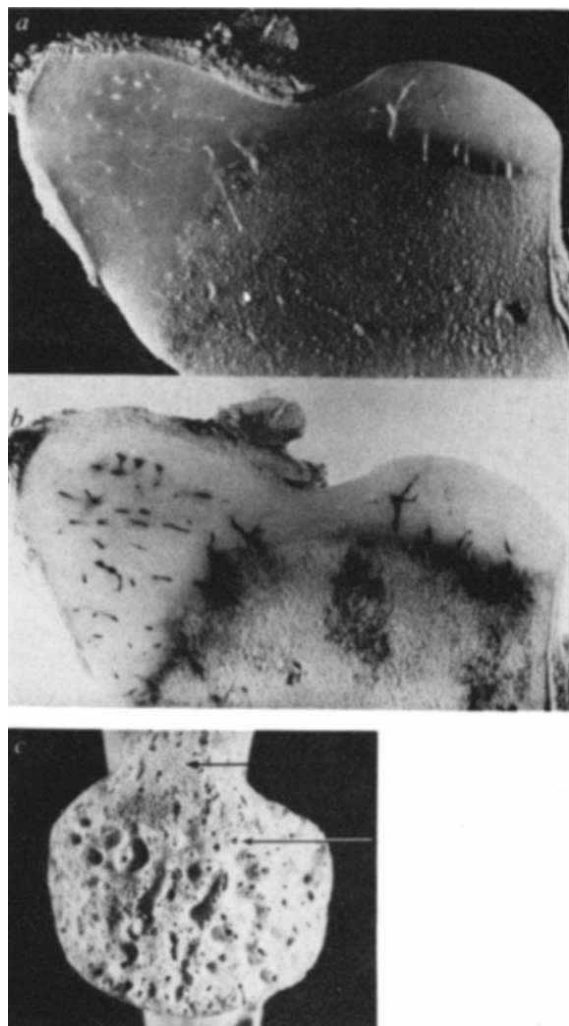


Fig. 1 Dried whole humeri were prepared by dermestid beetle maceration from two *Dermochelys coriacea*, one *Eretmochelys imbricata* (hawksbill), three *Chelonia mydas* (green turtle), two *Caretta caretta* (loggerhead) and two *Lepidochelys kempii* (ridley). Sections of whole formalin-fixed humeri, radii, ulnas and metacarpals were examined from two *C. caretta*, two *L. kempii* and four *D. coriacea* (straight-line carapace lengths 135, 153, 175 and 178 cm, estimated weights 260, 360, 470 and 470 kg; specimens found stranded in New Jersey during August–October 1979, and obtained from the Marine Mammal Stranding Center, Atlantic City). *a, b*, Natural and dye-injected longitudinal sections of the proximal humerus in a *D. coriacea* of 135-cm carapace length: medial process apophysis on the left, proximal epiphysis on the right. Small vessels of perichondral origin are seen in the centre of the apophysis; large transphyseal vessels are seen in both the epiphysis and the apophysis. *c*, View end-on of dried proximal humeral epiphyseal subchondral plate (cartilage removed) in a *D. coriacea* of ~190-cm carapace length. Arrowhead points to small metaphyseal end-arterioles participating in endochondral osteogenesis; black arrow points to large transphyseal fenestrations allowing vascular access to the centre of the cartilaginous epiphysis. Also seen are coarse growth plate undulations, which probably serve as stabilization for the large chondroepiphysis.

dark pigmentation in the periosteally derived trabecular bone and lack of internal ontogenetic remodelling (Fig. 2). Cones of light-coloured endochondral and dark-coloured periosteal bone are therefore formed, which radiate in a well delineated manner from the central nutrient artery located at the anatomic site of the initial diaphyseal ossification centre (see radiograph, Fig. 2).

Transverse sections of the mid-humerus show a cyclical variation in bone deposition which is mainly peripheral, both in the endochondral and periosteal portions. Other than pigmentation differences it is difficult to delineate grossly endochondral and

periosteal bone and the impression gained is of an amedullary cancellous region merging gradually peripherally into a slightly denser compacta, which is extensively vascular in the deeper portions but less so superficially, where it becomes a very thin shell of dense subperiosteal bone.

In reconstructing the hypothetical growth of this bone, it seems that both patterns of bone formation occur contiguously for a certain period of time, with the endochondral cones of bone progressively diverging from each other at their bases and the intervening space being filled in by periosteal bone. However, in the sub-adult animal it seems that further elongation of the bone becomes solely derived from endochondral growth, with very little contribution from the periosteal bone apart from the continued formation of a very thin cortical compacta. This apparent change in growth pattern may represent a relative onset of skeletal maturity comparable with closure of the physis in mammals. However, as secondary epiphyseal ossification and growth-plate fusion do not occur in the leatherback, it is capable of continual growth throughout adulthood, this apparently being mainly endochondral.

In all these features of chondro-osseous morphology the leatherback turtle appears most similar to marine mammals, notably the Cetacea (whales)¹¹ and Sirenia (manatees)¹². Both of these groups also vascularize their epiphyses perichondrally and transphyseally, develop amedullary bones with retained endochondral and periosteal cones that are pigmented and do not remodel, show a gradual merging of cancellous into compact bone, and in some forms (the migratory beluga whale) also show cyclical bone deposition^{11,12}. The only discernible differences in the leatherback are the lack of secondary epiphyseal calcification and ossification, the point of relative cessation of periosteal growth in the sub-adult animal, and the presence of dark pigmentation in the periosteal rather than the endochondral bone, which is a reversal of the pattern seen in the marine mammals.

No other known extant reptile shows this combination of chondro-osseous developmental features. Whereas some extant reptiles (for example, the varanid lizards, including the large Komodo dragon) vascularize their chondroepiphyses, this vascularization is always perichondrally or circumphyseally derived, never transphyseally¹³. In addition, the vascularization is invariably followed by secondary calcification and ossification of the epiphysis. All other turtles and all crocodilians previously examined, including species of large body size, fail to vascularize their thin chondroepiphyses, which also remain cartilaginous throughout adulthood^{8,13}. However, certain extinct reptiles show some similarities to the leatherback: plesiosaurs have been described as having endochondral and periosteal cones that do not remodel¹⁴; ichthyosaurs, plesiosaurs, nothosaurs and mesosaurs are known to have amedullary bones with cancellous-compacta differentiation paralleling that of marine mammals^{15,16}, and protostegid turtles have also been noted briefly to be somewhat similar in this respect¹⁷. All these fossil reptiles were highly adapted to a marine existence as shown by other skeletal features. Some fossil amphibians with pronounced marine modifications (certain stegocephalian species) also show a pattern of bone histology very similar to that described for marine mammals¹⁸. Other extant vertebrates with marine adaptations have similar developmental patterns, notably pinnipeds (seals)¹⁹ and penguins²⁰. However, some marine vertebrates do not show this spectrum of characteristics—notably the typical hard-shelled marine turtles (Cheloniidae) which are, in most respects, similar to freshwater or terrestrial species.

No terrestrial vertebrate has these chondro-osseous developmental features. The fact that such diverse groups as cetaceans, sirenians, pinnipeds, penguins, extinct marine reptiles and amphibians, and leatherback turtles have such a high degree of physical similarity in bone morphology suggests an underlying mechanism of marine adaptability which has led to a highly developed pattern of skeletal evolutionary convergence. The failure of some marine vertebrates (for example, hard-shelled turtles) to develop these characteristics fully may

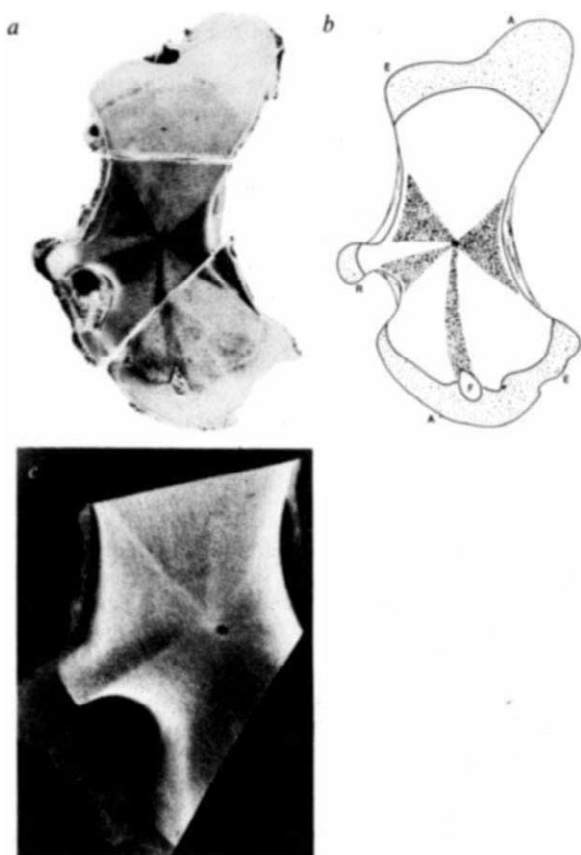


Fig. 2 *a, b*, Actual and schematic longitudinal sections of unstained humerus of a *D. coriacea* of 135-cm carapace length. Endochondral bone is light coloured whereas membranous (periosteal) bone is dark. Note also the amedullarity, lack of clear cancellous-compacta differentiation and disappearance of pigmentation at the superficial cortical margin. E, epiphyses; A, apophyses; F, ectepicondylar foramen; R, radial process apophysis. *c*, Radiograph of thin longitudinal slice of central diaphyseal portion of humerus showing central nutrient foramen and radiating cones of endochondral and membranous bone. Orientation as in Fig. 2a.

be related either to an incompletely evolved system due to relatively recent marine adaptation, or to a different set of physiological demands being required for differing life habits. Some studies have suggested that the marine pattern of bone development may be related to the homeostatic requirements of prolonged deep diving^{19,21,22}. It has been hypothesized that the intense dysbaric and acidotic challenges produced by repetitive deep diving can be countered by a skeletal buffering system which becomes more effective through increased vascularity in denser, amedullary bone¹⁹. Probably also of significance are the altered skeletal biomechanics of a marine existence where the limbs are not needed to support weight but instead are subject to entirely different forces produced by propulsion or steering in a medium much denser than air. Theories have also been advanced supporting hypothyroidism¹² and altered thermal relationships¹¹ as the underlying mechanisms leading to marine skeletal characteristics.

With regard to the phylogeny of *Dermochelys*, theories have vacillated between those supporting extreme specialization in a recently derived form and those favouring extreme primitiveness. Current taxonomic opinion accepts the former theory and is based on many similarities to the hard-shelled turtles²³⁻²⁸. However, strong arguments supporting the latter have also been advanced in the past^{7,8,29,30}. We suggest that the chondro-osseous morphology of *Dermochelys* emphasizes its extreme distinctness from other extant marine turtles (and turtles and reptiles in general). Whether this distinctness is recently derived or primitive cannot yet be answered.

Theories concerned with the phylogeny of endothermy and the origin of mammals from reptiles have stressed the relationships of increased body size and thermal inertia³¹ as well as altered bone histology³². It has been suggested that endothermy evolved separately in a number of extinct reptile groups³¹, and that transphyseal vascularization of a permanently cartilaginous epiphysis may have been the pattern in early mammal-like reptiles³³. The striking similarities in bone growth between *Dermochelys* and certain marine mammals, the presence of transphyseal vessels without secondary epiphyseal ossification, and the well developed homoiothermic (or endothermic) thermoregulatory adaptations suggest that this large animal may have developed along a course parallel to certain extinct endothermic reptiles, although in a group not previously thought to be undergoing this evolutionary process. In addition, the leatherback turtle's chondro-osseous morphology represents a striking and seemingly unique parallel between an extant reptile and modern marine mammals.

We thank the following people for their help and comments: R. C. Schoelkopf, A. Mead, J. P. Ross, E. E. Williams, H. C. W. Skinner and C. R. Shoop. This work was supported in part by grants R01-HD-10854 and K04-AM-00300 from the NIH.

Received 18 September 1980; accepted 21 January 1981.

1. Pritchard, P. C. H. *Catalog of American Amphibians and Reptiles*, 238 (1980).
2. Frair, W., Ackman, R. G. & Mrosovsky, N. *Science* **177**, 791 (1972).
3. Greer, A. E., Lazell, J. D. Jr & Wright, R. M. *Nature* **244**, 181 (1973).
4. Neill, W. H. *et al. Science* **184**, 1008 (1974).
5. McNab, B. K. & Auffenberg, W. *Comp. Biochem. Physiol.* **55A**, 345 (1976).
6. Gervais, P. *Mem. Nouv. Arch. Mus. nat. Hist.*, Paris **8**, 199 (1872).
7. Dollo, L. *Bull. Mus. R. Hist. nat. Belg.* **5**, 59 (1888).
8. Deraniyagala, P. E. P. *Proc. zool. Soc. Lond.* **68**, 1057 (1930).
9. Haines, R. W. *J. Anat.* **72**, 323 (1938).
10. Suzuki, H. K. *Ann. N.Y. Acad. Sci.* **109**, 351 (1963).
11. Felts, W. J. I. & Spurrell, F. A. *Am. J. Anat.* **118**, 103 (1966).
12. Fawcett, D. W. *Am. J. Anat.* **71**, 271 (1942).
13. Haines, R. W. *J. Anat.* **75**, 282 (1941); *Biol. Rev.* **17**, 267 (1942); in *Biology of the Reptilia* Vol. 1 (eds Gans, C., Bellairs, A. d'A. & Parsons, T. S.) 81-115 (Academic, London, 1969).
14. Moodie, R. L. *Am. J. Anat.* **7**, 443 (1908).
15. Enlow, D. H. & Brown, S. O. *Tex. J. Sci.* **9**, 186 (1957).
16. Ricqlès, A. J. de in *Morphology and Biology of Reptiles* (eds Bellairs, A. d'A. & Cox, C. B.) 123-150 (Academic, London, 1976).
17. Cope, E. D. *Proc. Am. phil. Soc.* **12**, 422 (1872).
18. Ricqlès, A. J. de *Colloques int. Cent. nat. Rech. Sci.* **218**, 351 (1975).
19. Versaggi, C. S. *Sci. Biol. J.* **3**, 305 (1977).
20. Meister, W. *Anat. Rec.* **143**, 377 (1962).
21. Nopcsa, F. *Anat. Anz.* **56**, 353 (1923).
22. Sickenberg, O. *Palaeobiologica* **4**, 405 (1931).
23. Case, E. C. *J. Morph.* **14**, 21 (1897).
24. Wieland, G. R. *Am. J. Sci.* **9**, 413 (1900); **14**, 95 (1902); **27**, 101 (1909).
25. Romer, A. S. *Osteology of the Reptiles* (University of Chicago Press, Chicago, 1956).
26. Gaffney, E. S. *Bull. Am. Mus. nat. Hist.* **155**, 387 (1975).
27. Wermuth, H. & Mertens, R. *Tierreich* **100**, 1 (1977).
28. Frair, W. *Herpetologica* **35**, 239 (1979).
29. Hay, O. P. *Am. Nat.* **32**, 929 (1898).
30. Zangerl, R. *Fieldiana, Geol. Mem.* **3**, 57 (1953).
31. McNab, B. K. *Am. Nat.* **112**, 1 (1978).
32. Ricqlès, A. J. de *C. r. hebdom. Séanc. Acad. Sci., Paris* **268D**, 782 (1969); *Evol. Theory* **1**, 51 (1974); in *Morphology and Biology of Reptiles* (eds Bellairs, A. d'A. & Cox, C. B.) 123-150 (Academic, London, 1976).
33. Ricqlès, A. J. de *Année biol.* **18**, 1 (1979).

Genetic control for sibling recognition?

Andrew R. Blaustein & Richard K. O'Hara

Department of Zoology, Oregon State University, Corvallis, Oregon 97331, USA

The ability to distinguish between kin and non-kin is critical to current theories of altruistic behaviour and kin selection¹. Hamilton¹ predicted that individuals would behave differently towards one another depending on the genetic relatedness between them. When either proximity to or familiarity with kin is a good predictor of relatedness, the mechanism by which favouritism towards kin is accomplished may not require special kin recognition abilities^{2,3}. However, if proximity and familiarity are poor predictors of kinship, favouritism (hence increases in inclusive fitness⁴) could only be achieved by the differential

recognition of kin and non-kin. We have previously shown that *Cascades frog* (*Rana cascadae*) tadpoles reared with siblings prefer to associate with siblings over non-siblings⁵. The present study is the first to report that totally naive individuals (*R. cascadae* tadpoles) prefer to associate with siblings over non-siblings. Because tadpoles were separated before hatching and reared apart from other individuals, results suggest that the ability of these tadpoles to discriminate between siblings and non-siblings has some innate component.

Sibling recognition has been reported in experimental studies of spiny mice (*Acomys cahirinus*)^{6,7}, pigtail macaques (*Macaca nemestrina*)⁸ and American toad tadpoles (*Bufo americanus*)⁹. In each case, young were separated after birth or hatching and later given a choice of associating with siblings and non-siblings. In the mouse studies, sibling recognition was probably a result of learning during early development^{6,7}. Sibling recognition in pigtail macaques is possible in the absence of prior sibling association. It is unclear how sibling recognition develops in toad tadpoles. Although the existence of a genetic recognition system has been proposed in theoretical discussions of kin selection^{1,10,11}, experimental evidence (with the possible exception of the macaque study⁸) is lacking.

Two *R. cascadae* clutches were collected in the Oregon Cascade Mountains and transported to the laboratory in separate containers. Twenty eggs from one clutch (group A) were placed individually in opaque containers (A_1 individuals) at early embryonic development (yolk plug stage; stage 12, ref. 12). The remainder of the clutch (250 tadpoles) was reared in a single aquarium (group A). A second clutch was divided in half (groups B_1 and B_2 ; developmental stage 19, ref. 12) and each group was placed in a separate aquarium.

A rectangular tank (122 × 44 × 30 cm) was used to test two tadpoles simultaneously for sib group preferences. To create end compartments, a partition of 1.5-mm plastic mesh was placed 15 cm from each end of the tank. The remaining central portion of the tank was longitudinally divided by an opaque, water-tight partition. When testing A_1 individuals, 25 tadpoles from group A and 25 tadpoles from group B_1 were placed in opposite ends of the tank. When testing B_1 individuals, 25 tadpoles from groups A and B_2 were placed in opposite ends of the tank. For each test, these stimulus tadpoles were left undisturbed for 15 min. Then, at the tank centre, one tadpole was released on each side of the partition and allowed to acclimatize for 10 min. Each tadpole was tested in four 5-min trials at 10-min intervals. We recorded time spent in seconds in sib and non-sib halves of the tank.

Tadpoles reared in total isolation since early embryonic development (A_1) and tadpoles reared with sibs (B_1) but given a choice of unfamiliar sibs or non-sibs preferred to associate with sibs. Both the number of individuals spending most of their time on the sib side of the tank (18/20 for A_1 and 16/20 for B_1 ; $P = 0.001$ and $P = 0.006$ respectively, binomial test) and the number of seconds spent by each of these individuals were significant (Table 1). These results provide the first experimental evidence that an organism can recognize siblings after being separated from sibs while still embryos and after being raised in complete isolation.

Long-term field observations on the social and ecological conditions surrounding early development of *R. cascadae* indicate that ample opportunity exists for sibling groups to develop¹³. Females oviposit eggs in specific habitats in one or more localities around a lake or pond. Clutches are usually laid next to or on top of other conspecific egg masses. As they develop, dispersal from the site of oviposition tends to be low. Furthermore, larvae were never observed singly but always in close association with conspecifics. This tendency for *R. cascadae* to aggregate has been verified experimentally¹³. It is likely that natural groupings of this species are comprised of a high proportion of siblings and early development occurs in the presence of sibs and non-sibs.

Predation is probably one selective force for group formation in *R. cascadae* as they are palatable and preyed on by several

Table 1 Mean times out of 1,200 s spent by individuals on sib side of test tank when given a choice between sibs and non-sibs

Individuals tested	n	Seconds spent on sib side (mean ± s.e.m.)	T
A_1	20	702.6 ± 20.65	15*
B_1	20	682.2 ± 33.22	50†

Tests of A_1 tadpoles and B_1 tadpoles were made at 14–18 and 19–23 days post-hatching respectively. The Wilcoxon signed rank test (one-tailed; t represents the test statistic)¹⁸ was used to test for significant departures from an expected random distribution of total times on sib and non-sib ends. Individuals from group A_1 were reared in 0.5-l containers. The other members of groups A, B_1 and B_2 were reared in 38-l aquaria. All tadpoles were reared in the same room in a 14-h light 10-h dark photoperiod. The temperature in the laboratory was kept at 20–22 °C. Tadpoles were fed rabbit pellets *ad libitum*. The same 25 stimulus individuals were used for five tests, and were then discarded. Between tests, end compartment placements of As and Bs were alternated. Previously, two controls (20 tadpoles each from two different clutches) were run in the absence of stimulus groups to determine possible biases in the testing apparatus. In these controls tadpoles displayed a random distribution pattern⁵. Following each test, tadpoles were removed, the tank drained, rinsed and refilled to a depth of 6 cm with 34 l of dechlorinated tap water. No tadpole was tested more than once. Average developmental stages¹² for groups A_1 , A and B were 28.5, 28.0 and 28.0 respectively.

* $P < 0.005$; † $P < 0.025$.

species^{13,14}. Individuals can increase their fitness by living in groups and detection of predators increases with increasing group size¹⁵. Although kinship alone is not a necessary prerequisite for the evolution of groups, an individual in a group consisting of many sibs who warns others of a predator may increase its inclusive as well as its individual fitness compared with an individual giving a similar warning in a group with fewer or no sibs^{16,17}.

Our data may be added to the few reports^{6–9} which suggest that some species preferentially associate with sibs over non-sibs. Significantly, there is a major similarity and a major difference between the present study and that of Wu *et al.*⁸ which greatly enhance the conclusions of both studies. The use of paternal half sibs by Wu *et al.*⁸ and our use of individuals reared in isolation from an early embryonic stage show that individuals can recognize sibs without prior sib association. However, the macaques used by Wu *et al.*⁸ were allowed to associate with non-sib conspecifics during development whereas tadpoles in our study were reared in total isolation and had no exposure to conspecifics during development. Furthermore, we suggest that the ability to recognize sibs in *R. cascadae* may exist in early embryonic life. The results of this and another study to be reported in detail elsewhere⁵ are consistent with a genetic recognition system. In the latter study, two different clutches were reared from an embryonic stage in an aquarium separated by plastic mesh. This allowed chemical and visual contact between tadpoles during development. When given a choice of unfamiliar sibs or unfamiliar non-sibs, these tadpoles associated with unfamiliar sibs. When given a choice between a group comprised of an equal mixture of sibs and non-sibs (tadpoles with which they were reared) and an unfamiliar sib group, tadpoles associated with unfamiliar sibs. Although it is possible that individually reared tadpoles could imprint on their own cues (or subtle maternal cues from the egg mass jelly), results of tadpoles reared with non-sibs⁵ argue against this.

In organisms that undergo early development in the presence of non-kin (such as *R. cascadae*) or in those that disperse before associating with kin, a kin recognition mechanism that functions independently of experience may be required to achieve recognition. Our experiments are consistent with such a mechanism.

We thank P. S. Dawson, W. G. Holmes, R. H. Porter, W. R. Rice, P. B. Samollow, R. R. Warner and J. A. Wiens for valuable comments. M. Bekoff, R. H. Porter and W. R. Rice critically reviewed the manuscript.

Received 3 October 1980; accepted 13 January 1981.

1. Hamilton, W. D. J. *theor. Biol.* **7**, 1–52 (1964).
2. Sherman, P. W. in *Sociobiology: Beyond Nature/Nurture* (eds Barlow, G. W. & Silverberg, J.) 505–544 (Westview, Boulder, 1980).
3. Bekoff, M. in *Parental Care in Mammals* (eds Gubernick, D. & Klopfer, P. H.) 307–346 (Plenum, New York, 1981).
4. Wilson, E. O. *Sociobiology: The New Synthesis* (Harvard University Press, 1975).
5. O'Hara, R. K. & Blaustein, A. R. *Anim. Behav.* (submitted).
6. Porter, R. H., Wyrick, M. & Pankey, J. *Behav. Ecol. Sociobiol.* **3**, 61–68 (1978).
7. Porter, R. H. & Wyrick, M. *Anim. Behav.* **27**, 761–766 (1979).
8. Wu, H. M. H., Holmes, W. G., Medina, S. R. & Sackett, C. P. *Nature* **285**, 225–227 (1980).
9. Waldman, B. & Adler, K. *Nature* **282**, 611–613 (1979).
10. Dawkins, R. *The Selfish Gene* (Oxford University Press, 1976).
11. Rothstein, S. I. *J. theor. Biol.* **87**, 255–261 (1980).
12. Gosner, K. L. *Herpetologica* **16**, 183–190 (1960).
13. O'Hara, R. K. thesis, Oregon State Univ. (1981).
14. Briggs, J. L. & Storm, R. M. *Herpetologica* **26**, 283–300 (1970).
15. Bertram, B. C. R. in *Behavioural Ecology: An Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 64–96 (Sinauer, Sunderland, 1978).
16. Sherman, P. W. *Am. Nat.* **113**, 925–935 (1979).
17. West-Eberhard, M. J. *Q. Rev. Biol.* **50**, 1–34 (1975).
18. Hollander, M. & Wolfe, D. A. *Nonparametric Statistical Methods* (Wiley, New York, 1973).

Substrate-adhering lymphoid cells show impaired tumorigenicity and increased immunogenicity

Jacob Hochman*, Aviva Katz, Efrat Levy & Shmuela Eshel

Department of Zoology, The Hebrew University of Jerusalem, Jerusalem, Israel

Cell adhesiveness is fundamental to a variety of biological phenomena, including tumour development and metastasis^{1–4}. Recently, Bubenik *et al.*⁵ have reported that in various malignant fibroblastoid cell lines those cells which demonstrated less adhesiveness were more tumorigenic. The relationship between cell adhesiveness, transformation and metastasis has been studied extensively in cells (fibroblastoid) grown as monolayers attached to their substratum and to each other^{4,6–9}, but, to our knowledge, there has been no report describing this relationship in suspension-borne (lymphoid) cells that grow free of each other and their substratum. We report here that substrate-adhering variant cells, selected from the tumorigenic, suspension-growing S49 mouse lymphoma, have impaired tumorigenicity. Furthermore, the substrate-adhering cells also have increased immunogenicity, as their inoculation into mice protects the mice from subsequent challenges with parental, non-adherent tumorigenic S49 cells. These findings suggest a new approach for the selection of immunogenic cells from suspension-borne tumorigenic cell populations.

In our study, a subline of S49 mouse lymphoma cells¹⁰, descendants of clone 24.6.1 (ref. 11), was propagated in suspension culture and readapted to grow as a transplantable tumour in BALB/c mice¹². Selection for cells with increased tumorigenicity was carried out through serial passages as tumours in BALB/c mice. Tumour cells from the 25th passage, which formed tumours as early as 7–10 days post-inoculation, were transferred into culture medium (subline T-25). These cells retained the same tumorigenicity *in vivo* for at least 6 months after transfer into suspension culture. From T-25 cells we have selected without the use of mutagens a subline (designated T-25-Adh) able to adhere to the substrate bottom of tissue culture flasks (Table 1). This selection involved continuous enrichment of substrate-adhering cells by discarding the cells growing in suspension. While attached to their substrate, these cells continued to grow until a dense monolayer was formed.

Table 1 Comparison between parental T-25 cells and their adherent variant (T-25-Adh) subline

Cells	Substrate adhesiveness	Tumorigenicity*	Immunizing† ability against T-25 cells in syngeneic mice
T-25	–	+	–
T-25-Adh	+	–	+

Cells were propagated in culture in Dulbecco's modified Eagle's medium (Gibco), supplemented with 10% heat-inactivated horse serum, penicillin (50 U ml^{–1}) and streptomycin (50 µg ml^{–1}). Cells were maintained at 37 °C in a humidified atmosphere containing 5% CO₂. Adhering cells were detached from their substratum (bottom of tissue culture flask) by shaking. Cells were adjusted to 1 × 10⁶ cells ml^{–1} and 10-ml portions were added to 50-ml tissue culture flasks (Costar 3050, 25 cm²) and incubated at 37 °C. At various intervals thereafter, flasks were tilted, aliquots taken out and the number of cells remaining unattached scored. From this, the percentage of attached cells was calculated. Although none of the parental T-25 cells attached, essentially all T-25-Adh cells attached to the substrate within 45–60 min. Assays were carried out only on samples in which viability, according to the trypan blue exclusion test, was > 95%.

* BALB/c males, 8–10 weeks old, were inoculated with 1 × 10⁷ cells intraperitoneally. Mice were checked daily, and the first day on which tumours could be recognized was recorded. T-25 cells (*n* = 16) developed tumours within 7–10 days and all the mice died within 13–15 days. T-25-Adh cells (*n* = 14) did not develop any tumours for at least 90 days.

† BALB/c males inoculated intraperitoneally 90 days earlier with 1 × 10⁷ T-25-Adh cells (*n* = 10) and control BALB/c males (*n* = 4) were both inoculated with tumorigenic T-25 cells (1 × 10⁷ cells per mouse, intraperitoneally). Control mice developed tumours and died within 15 days, whereas all the mice that had previously received T-25-Adh cells did not develop any visible tumours for at least 180 days.

When the substrate was fully covered by a monolayer, cells appeared in the liquid phase of the medium above. These cells then continued to divide normally as a suspension of single cells or small aggregates of a few cells. When the free-floating cells were transferred to a new culture flask, they attached to the substrate with the same kinetics as their parental adherent cells.

When T-25 and T-25-Adh cells were inoculated into BALB/c mice, only mice that had received the T-25 (non-adherent) cells developed tumours (100% within 10 days) and died within 13–15 days post-inoculation. On the other hand, mice that received the same inoculum of T-25-Adh (substrate-adhering) cells did not demonstrate any visible tumours for at least 90 days (Table 1). These findings demonstrate that the selection for substrate-adhering S49 cells significantly decreases their tumorigenicity in BALB/c mice.

We tested whether the decrease in tumorigenicity of the substrate-adhering cells involved an immune response component, by exposing BALB/c mice that had previously received T-25-Adh cells to a challenge with 10⁷ tumorigenic T-25 cells per mouse. BALB/c mice of the same age group that had not been previously subjected to T-25-Adh cells were controls (Table 1). T-25-Adh cells were able to confer on BALB/c mice immunity against a challenge of suspension-growing tumorigenic T-25 cells. Furthermore, in preliminary experiments, the immunizing ability against T-25 cells could be adoptively transferred to BALB/c mice via spleen cells from mice previously inoculated with adherent (T-25-Adh) cells.

Both the selection for substrate adhesiveness and immunization procedures have been repeated with another tumorigenic subline, T-40 (derived from the 40th passage *in vivo*), and its selected adherent variant, and a tumorigenic subline derived from descendants of a different clone (24.3.2) of S49 cells¹¹. Here again, the adherent variant cells demonstrated impaired tumorigenicity concomitantly with increased immunogenicity when compared with their suspension-growing parental cells.

* Present address: Laboratory of Molecular Biology, National Cancer Institute, NIH, Bethesda, Maryland 20205, USA.

This report demonstrates for the first time that the relation between increased cell adhesiveness and decreased tumorigenicity^{5,9,13} may be applied to malignant cells of the lymphoid category. Furthermore, the impaired tumorigenicity of substrate-adherent S49 cells is, at least in part, the result of an increase in their immunizing ability.

Xenogenization or heterogenization are general terms now used to describe all attempts to make tumour cells antigenically foreign to their host so that they may be used more effectively in immunotherapy and immunodiagnosis. Several approaches are used for the biological xenogenization of tumour cells¹⁴⁻²¹, a comprehensive summary of which was published recently²². Our findings in S49 cells suggest a new and somewhat different approach for the xenogenization of malignant lymphoid cells. Here, the isolation of immunogenic, non-tumorigenic cells is based on a simple selection for substrate adherence. The nature of the cell-surface change(s) that took place on transition from free-floating to substrate-adhering, non-tumorigenic, immunogenic S49 cells is being investigated. Furthermore, it is of both mechanistic and practical significance to find out whether the approach taken here is also applicable to other lymphoid tumour cell lines.

This work was supported by a special fund from the faculty of Mathematics and Natural Sciences, The Hebrew University of Jerusalem, to J. H.

Received 26 September 1980; accepted 7 January 1981.

1. Ludford, R. J. *Proc. R. Soc. B* **112**, 250-263 (1932).
2. Cowdry, E. V. *Archs Path.* **30**, 1245-1274 (1940).
3. Coman, D. R. *Cancer Res.* **4**, 625-647 (1944).
4. Grinnel, F. *Int. Rev. Cytol.* **53**, 65-143 (1978).
5. Bubenik, J. et al. *Int. J. Cancer* **23**, 392-395 (1979).
6. Pouyssegur, J. M. & Pastan, I. *Proc. natn. Acad. Sci. U.S.A.* **73**, 544-548 (1976).
7. Winkelhake, J. L. & Nicolson, G. L. *J. natn. Cancer Inst.* **56**, 285-291 (1976).
8. Briles, E. B. & Kornfeld, S. *J. natn. Cancer Inst.* **60**, 1217-1222 (1978).
9. Murray, J. C., Liotta, L., Rennard, S. I. & Martin, G. R. *Cancer Res.* **40**, 347-351 (1980).
10. Horibata, K. & Harris, A. W. *Expl Cell Res.* **60**, 61-77 (1970).
11. Coffino, P., Bourne, H. R. & Tomkins, G. M. *J. cell. Physiol.* **85**, 603-610 (1975).
12. Hochman, J., Katz, A., Levy, E. & Eshel, S. (in preparation).
13. Coman, D. R. *Cancer Res.* **14**, 519-531 (1954).
14. Lindenmann, J. *J. Immun.* **92**, 912-919 (1964).
15. Lindenmann, J. *Biochim. biophys. Acta* **355**, 49-75 (1974).
16. Bekesi, G., Arneault, S. & Holland, J. *Proc. Am. Ass. Cancer Res.* **12**, 47 (1971).
17. Hashimoto, Y. & Yamanoha, B. *Gann* **67**, 315-319 (1976).
18. Hashimoto, Y., Yamanoha, B., Endoh, H. & Ishizuka, T. in *Gann Monograph on Cancer Research* (ed. Kobayashi, H.) **23**, 135-141 (Japan Scientific Societies, Tokyo, 1979).
19. Harris, H., Miller, O. L., Klein, G., Worst, P. & Tachibana, T. *Nature* **223**, 363-366 (1969).
20. Klein, G. & Harris, H. *Nature new Biol.* **237**, 163-164 (1972).
21. Shintzky, M., Skornick, Y. & Haran-Ghera, N. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5313-5316 (1979).
22. *Gann Monograph on Cancer Research* (ed. Kobayashi, H.) **23** (Japan Scientific Societies, Tokyo, 1979).

Fibroblast traction as a mechanism for collagen morphogenesis

Albert K. Harris, David Stopak & Patricia Wild

Department of Zoology, University of North Carolina, Wilson Hall 046-A, Chapel Hill, North Carolina 27514, USA

To make visible the traction forces exerted by individual cells, we have previously developed a method of culturing them on thin distortable sheets of silicone rubber¹. We have now used this method to compare the forces exerted by various differentiated cell types and have examined the effects of cellular traction on re-precipitated collagen matrices. We find that the strength of cellular traction differs greatly between cell types and this traction is paradoxically weakest in the most mobile and invasive cells (leukocytes and nerve growth cones). Untransformed fibroblasts exert forces very much larger than those actually needed for locomotion. This strong traction distorts collagen gels dramatically, creating patterns similar to tendons and organ capsules. We propose that this morphogenetic rearrangement of extracellular matrices is the primary function of fibroblast traction and explains its excessive strength.

As a tissue culture cell spreads on one of our silicone rubber substrata, the traction force with which the cell margin pulls itself forwards compresses an area of the rubber sheet beneath the cell (Fig. 1a). The thin rubber sheet becomes folded into wrinkles of two kinds: compression wrinkles directly beneath the cell, running transverse to the force, and tension wrinkles which radiate outwards from the cell margins in the direction of maximum stretching. Of the cells we have examined, the strongest traction was exerted by fibroblasts from embryonic chick heart, skeletal muscle and tendon, dermis and the choroid coat of the eye, and by BALB/3T3 mouse cells. Individual cells of these types produced numerous wrinkles beneath and around themselves (Fig. 1a, b). Glial cells from dorsal root ganglia and platelets from human blood also exerted very strong traction. In contrast, epithelial cells (from embryonic chick epidermis, pigmented retina and liver parenchyma) were considerably weaker. These could still distort the rubber sheet, but only when many

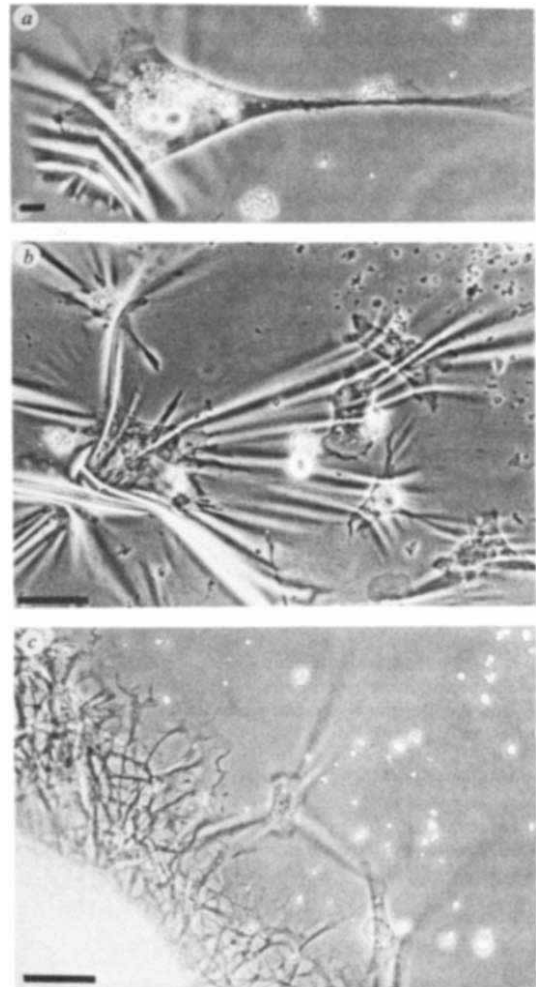


Fig. 1 Tissue culture cell spreading on thin layers of silicone rubber. The traction by which the cells spread and propel themselves is shown by the wrinkles formed in the rubber sheet. Silicone rubber substrata were made by flaming the surface of 30,000 cP silicone fluid (polydimethylsiloxane, Dow-Corning; see ref. 1 for details of method). *a*, An individual 3T3 cell which has compressed the rubber layer beneath its leading margin into numerous folds. We call this pattern of distortion compression wrinkles. Scale bar, 10 μ m. *b*, Group of dissociated chick heart fibroblasts produce a more complex pattern of substrate distortion, composed of compression wrinkles and tension wrinkles, which radiate from the cell margins and often run between cells. Scale bar, 100 μ m. *c*, Although numerous growth cones extend in a wide band from this explanted dorsal root ganglion, no wrinkles form beneath or adjacent to them. The wrinkles visible are tension wrinkles originating from the two peripheral glial cells. Scale bar, 100 μ m.

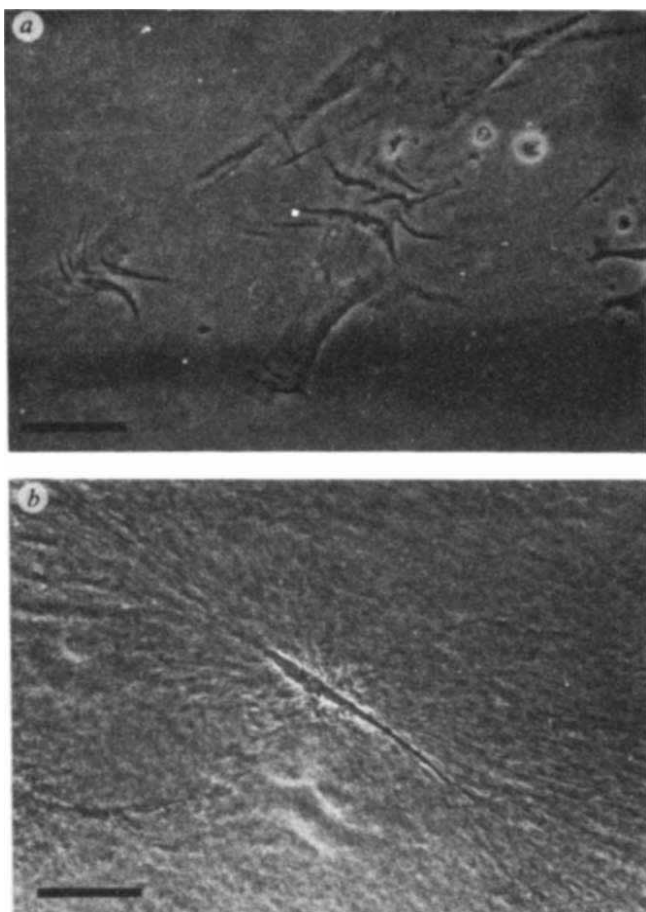


Fig. 2 Collagen substrata are distorted by cellular traction in a pattern similar to the wrinkling of silicone rubber. *a*, Dermal fibroblasts spreading over a thin film of dried collagen produce a pattern of substrate folding, equivalent to compression wrinkling. Collagen solution (Vitrogen 100, Collagen Corporation) was pipetted onto glass coverslips, dried with a stream of sterile air and washed with phosphate-buffered saline (PBS) to remove the acid. Trypsin-dissociated fibroblasts from 7-day chick embryo skin were then seeded onto the collagen films. *b*, Dermal fibroblasts grown in reconstituted collagen gels do not generate wrinkles, but do align the flexible collagen meshwork around themselves. Around the cell body collagen is packed perpendicular to the direction of cell locomotion, while beyond the cell margins, sprays of collagen are aligned parallel to the long axis of the cell. This is equivalent to the compression and stretching seen in Fig. 1*b*. Collagen gels were prepared by mixing collagen solution with $5 \times$ Dulbecco's modified Eagle's medium, fetal calf serum, NaOH and PBS (5:2:1:0.7:1.3 by vol) at 4 °C. Bar, 50 μ m.

cells pulled in concert. Weaker still were transformed cells of the lines L-929 and CHO as well as SV40 virus-transformed 3T3 cells. These cells could only deform the rubber visibly where it was already wrinkled or torn. The weakest traction was exerted by macrophages (from mouse peritoneum), polymorphonuclear leukocytes (from human blood) and nerve growth cones (from embryonic chick sensory ganglia). Their weakness was not just a reflection of smaller size. Even when dozens of nerve growth cones lay side by side, their cumulative tractional strength was still too weak to distort our weakest rubber substrata (Fig. 1*c*).

Hence the paradox is that the most mobile cell types exert the weakest traction while the contractile strength of fibroblasts is very much larger than that needed for cell locomotion. To estimate the factor by which fibroblast traction exceeds that needed for locomotion, we have analysed time lapse films in which one margin of a spreading fibroblast breaks loose from the

wrinkled substratum so that it is pulled rapidly sideways by the elastic relaxation of the rubber. Such elastic retractions often displace fibroblasts by distances of many tens or even hundreds of micrometres, distances which correspond to the area of rubber sheet they had compressed beneath themselves while spreading. During these retractions the cell bodies are pulled at speeds of up to 10 μ m or more per second, which is much faster than the speed at which the margin of these cells advances during normal locomotion (0.5–1 μ m min⁻¹). When we consider that the elastic pull of the rubber cannot be any stronger than the cellular traction which originally distorted it, it becomes clear that the cell's traction is sufficient to propel it 100–1,000 times faster than normal cell locomotion. Thus we conclude that fibroblast traction is at least two or three orders of magnitude stronger than that needed to propel cells of this size at their normal speeds.

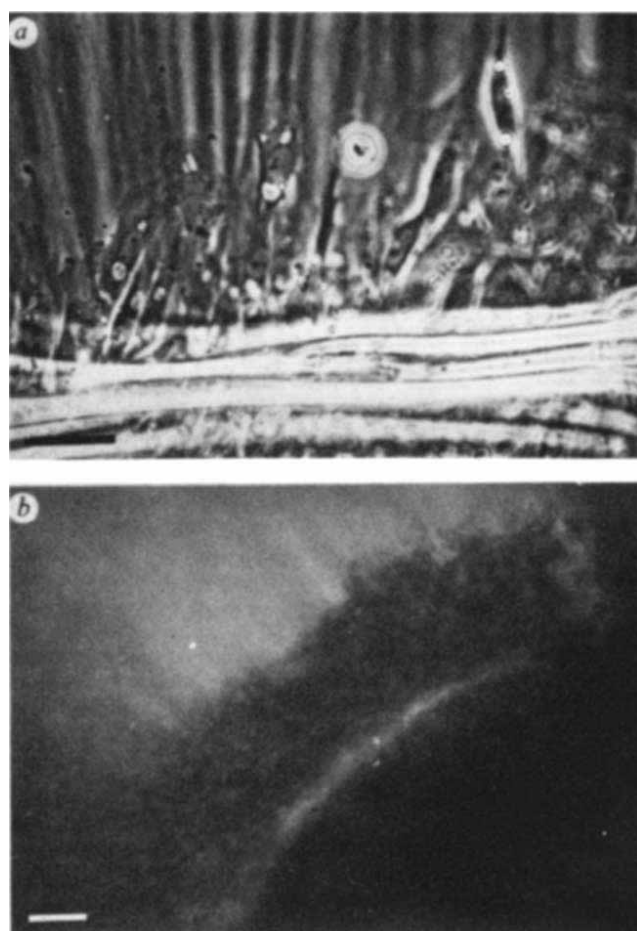


Fig. 3 Comparison of distorting effects of fibroblast traction on silicone rubber sheets and collagen gels. *a*, Marginal outgrowth zone of an explanted 8-day chick ventricle after 48 h in culture on silicone rubber. Adjacent to the explant (bottom of the field) the rubber layer is compressed into a mass of circumferential folds. Beyond the explant the silicone film is stretched resulting in radial wrinkles which continue far beyond the margins of the explant. *b*, The pattern of distortion is equivalent when heart explants are grown in collagen gels and viewed with polarized light. Adjacent to the explant a bright birefringent ring is visible consisting of a capsule of compressed collagen wrapped circumferentially around the explant. Next there is a dark band lacking birefringence where collagen is being compacted and pulled towards the explant by the migrating cells within this area. Collagen fibres farther from the explant are stretched, resulting in a radial pattern of fibre orientation. Scale bars, 100 μ m.

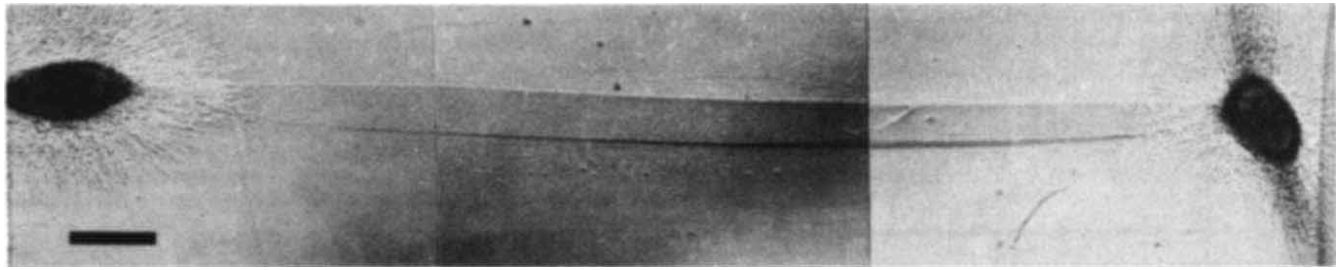


Fig. 4 Two centre effects form in collagen gels even when explants are separated by a distance spanning over 1.5 cm. Collagen fibres become aligned into long axially oriented tracts interconnecting two centres of traction. Heart explants from 8-day chick embryos after 96 h in culture. Scale bar, 1 mm.

Fibroblast traction ought to distort any sufficiently flexible material and we have studied its effects on collagen, the normal substratum of fibroblasts within the body². When aqueous solutions of collagen are air dried to form thin sheets, fibroblasts wrinkle these sheets just as they do rubber substrata (Fig. 2a). When the collagen is reprecipitated as a thick gel, wrinkles are no longer formed. Instead, fibroblast traction compresses and packs the collagen fibres together around and behind the cell bodies. Beyond the cellular extensions which exert this traction, the collagen fibres are stretched into alignment (Fig. 2b). Bell, Ivarsson and Merrill³ have previously described the contraction of collagen gels by fibroblasts. Our observations suggest that these distortions are caused by direct cellular traction—the combined motility and contractility of the cells exerting strong shear forces tangential to their surfaces. Note that this traction is distinct from simple contraction like that of a muscle. For example, the cells elongate instead of shorten as they compress and stretch the collagen around them.

When tissue fragments are explanted into collagen gels, the cumulative traction of many fibroblasts pulls the gel centripetally to form a dense capsule of compressed collagen fibres wrapped around the explant surface (Fig. 3). The region of gel farther from the explant is stretched into radial alignment with a 90° rotation in fibre alignment occurring where the fibroblasts exert their traction. Between adjacent explants the pull of this traction aligns the collagen into straight bundles (Fig. 4) which can be more than 1 cm long! The development of these traction fields demonstrates that collagenous structures such as tendons, ligaments, organ capsules and dermal tessellations⁴⁻⁶ need not be laid down by direct secretion of collagen fibres into their eventual arrangement⁷. Instead, fibroblast traction can rearrange and re-pack collagen into these patterns, even beginning from a totally random meshwork.

These traction fields resemble Paul Weiss' centre effects, which lends support to his theories of nerve guidance⁸ but suggests that guidance should be towards foci of strong traction rather than high growth rate. Our findings also point to fibroblast traction as the cause for the apparent contraction of collagen networks around deep wounds, burns and surgically implanted prostheses⁹⁻¹¹.

We had assumed originally that substratum distortion was merely an incidental by-product of cell locomotion, useful for studying its mechanism. However, these results suggest that traction has evolved to serve morphogenetic functions going well beyond cell displacement.

This work was supported by NIH grant GM-24251.

Received 16 October 1980; accepted 21 January 1981.

1. Harris, A. K., Wild, P. & Stopak, D. *Science* **208**, 177–179 (1980).
2. Elsdale, T. & Bard, J. *J. Cell Biol.* **54**, 626–637 (1972).
3. Bell, E., Ivarsson, B. & Merrill, C. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1274–1278 (1979).
4. Stuart, E. S. & Moscona, A. A. *Science* **157**, 947–948 (1967).
5. Wessels, N. K. & Evans, J. *Dev. Biol.* **18**, 42–61 (1968).
6. Weiss, P. A. in *Biological Organisation* (ed. Waddington, C. H.) 12–13 (Pergamon, New York, 1959).

7. Trelstad, R. L. & Hayashi, K. *Dev. Biol.* **71**, 228–242 (1979).
8. Weiss, P. A. in *Analysis of Development* (eds Willier, B. H., Weiss, P. A. & Hamburger, V.) 346–401 (Saunders, Philadelphia, 1955).
9. Peacock, E. E. & Van Winkle, W. *Wound Repair* Ch. 3 (Saunders, Philadelphia, 1976).
10. Abercrombie, M., Flint, M. H. & James, D. W. *J. Embryol. exp. Morph.* **4**, 167–174 (1954).
11. Gayou, R. & Rudolph, R. *Ann. Plastic Surg.* **2**, 62–71 (1979).

Genetically restricted cell-mediated cytotoxicity in cattle immune to *Theileria parva*

Elsie M. Eugui & David L. Emery

International Laboratory for Research on Animal Diseases,
PO Box 30709, Nairobi, Kenya

The protozoan parasite *Theileria parva*, which is transmitted by the tick *Rhipicephalus appendiculatus*, produces an acute fatal infection in the lymphoid system of susceptible cattle¹. This disease is a serious constraint to livestock improvement and production in large areas of East Africa¹. The parasite invades host lymphocytes, inducing rapid proliferation followed by widespread lymphocytolysis². Cattle which recover from theileriosis (East Coast fever) spontaneously, or which are immunized by infection and treatment with tetracycline³, are resistant to reinfection with the same isolate of *T. parva* for at least 3 yr⁴. Immunity against infection with the parasite cannot be ascribed to the production of specific antibodies⁵, but can be transferred adoptively between twins with thoracic duct leukocytes from the immunized partner⁶. These observations suggest that protective immunity is associated with cell-mediated mechanisms. We have now examined the capacity of leukocytes from immune cattle to lyse parasitized lymphoblastoid and non-parasitized tumour cell lines either directly or after stimulation in an autologous mixed lymphocyte reaction (MLR). In contrast to the nonspecific lytic activity of leukocytes from immune cattle reported by Pearson *et al.*⁷, we describe the sequential appearance in the lymph and blood of immune cattle, of cytotoxic leukocytes with activity restricted to target cells carrying the autologous genotype. These observations suggest that a major component of protective immunity to *T. parva* is mediated by cytotoxic cells which lyse parasitized cells in a genetically restricted fashion. We have also found that during a primary infection with *T. parva* cytotoxicity was manifested against allogeneic parasitized cells and xenogeneic uninfected target cells, but not against autologous infected cells. The features of cell-mediated immunity to *T. parva* during primary infection and immunization are discussed more fully elsewhere⁸.

Studies of cell-mediated immunity in cattle and other outbred species have been hampered by the lack of syngeneic cell lines for monitoring the development of cytotoxic responses. In the case of *T. parva*, this limitation has been overcome by using the

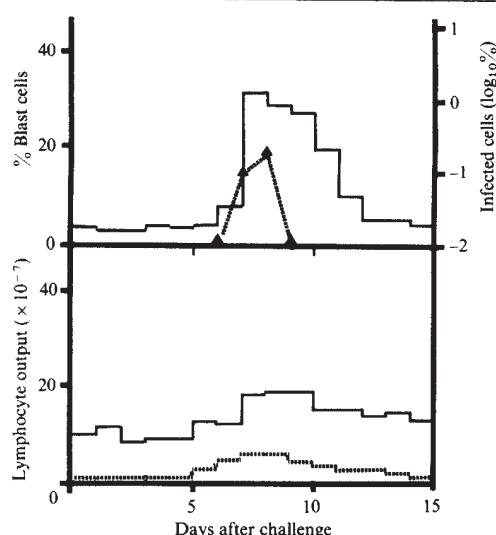


Fig. 1 The efflux in central lymph of normal and infected lymphocytes following challenge of three immune calves with fresh sporozoites adsorbed on autologous PBL. Results are expressed as mean average total (—) and blast lymphocyte output per hour (---); mean average percentage blast cells (—) and mean average percentage of lymphocytes infected with macrophages of *T. parva* (▲---▲), as determined by direct immunofluorescence¹¹.

ability of sporozoites of the parasite to transform *in vitro* leukocytes from normal cattle into lymphoblastoid cell lines infected with macrophages of *T. parva*⁹. We examined the specificity of cytotoxic responses of cattle immunized with *T. parva* using autologous and allogeneic infected target cells derived from such *in vitro* transformations, as well as xenogeneic non-parasitized cells (YAC-1). The latter is a mouse tumour cell line susceptible to lysis by natural killer (NK) cells of murine origin¹⁰, and we have found that it is lysed by normal bovine peripheral blood leukocytes (PBL).

Freisian calves aged 6–8 months, serologically negative for *T. parva*, were immunized by subcutaneous inoculation with sporozoites of *T. parva* (Muguga) and a concurrent intramuscular inoculation with long-acting tetracycline (20 mg per kg., Pfizer)³. Such immune animals were challenged subcutaneously 4 weeks later with the same isolate of the parasite from stablitate or as fresh sporozoites adsorbed on autologous PBL. At intervals after challenge, cytotoxic responses were monitored in PBL. After a further 6 weeks, three calves received a subcutaneous inoculum of 5×10^8 autologous infected cells, and two calves were inoculated with another dose of infective sporozoites. Again, cytotoxicity was evaluated in PBL at intervals after the second challenge.

Table 1 Specificity of the cytotoxic response in immune animals after challenge with 5×10^8 autologous *T. parva*-infected cells

PBL donor	Target cell, <i>Theileria</i> -infected	% Specific lysis* (days after challenge)	
256	Th-256	3	7
	Th-588	2.1	48.2
	Th-589	1.0	0
	Th-589	0	0
588†	Th-588	5.3	70.1
	Th-589	0	39.0
	Th-256	3.8	2.2
589†	Th-589	7.1	39.7
	Th-588	4.5	51.5
	Th-256	4.7	0
	Th-256	5.0	5.3
Normal‡	Th-588	5.2	4.3
	Th-589	2.3	0.5

* Ratio of effector to target cells was 100:1 in macroassays (Fig. 2 legend).

† Animals 588 and 589 are chimeric twins.

‡ Results are average of two normal control animals tested the same day as the immune challenged animals.

When immune animals were challenged with sporozoites as stablitate, macrophage-bearing cells were not detected in biopsies of the draining lymph nodes during the ensuing 3 weeks, nor in lymph effluent from these nodes. Lymph node cells (LNC) from the draining lymph nodes were unable to stimulate an autologous MLR using normal PBL cryopreserved before the primary infection, whereas they stimulated allogeneic

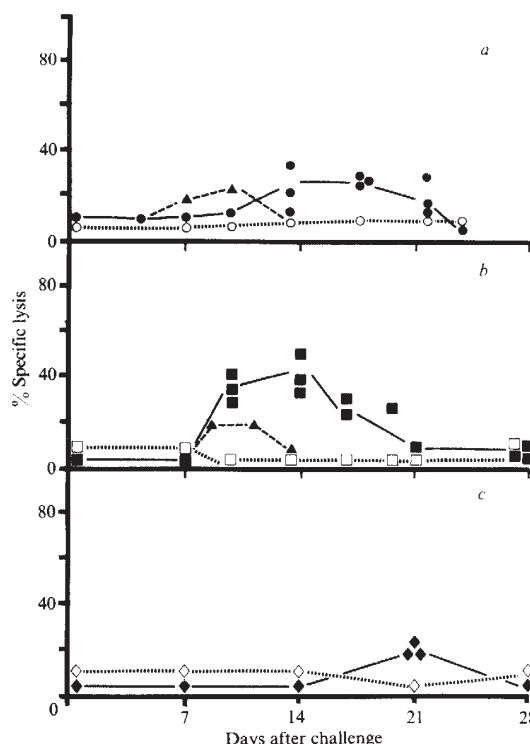


Fig. 2 Three calves were immunized as described and were inoculated subcutaneously 4 weeks later with a lethal dose of sporozoites of *T. parva* (Muguga) from stablitate. At intervals after challenge, PBL were purified using Ficoll-paque¹¹ (Pharmacia) and were tested in standard 4-h microcytotoxicity⁷ and macrocytotoxicity assays¹⁶ against autologous and allogeneic parasitized cell lines as well as a xenogeneic tumour cell line (YAC-1). All cells were suspended in RPMI-1640 culture medium (Flow Labs) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS), $20 \mu\text{g ml}^{-1}$ glutamine and 10 mM HEPES buffer. For the microcytotoxicity assay, 5×10^4 target cells (in 50 μl) which had been previously labelled with $200 \mu\text{Ci } ^{51}\text{Cr}$ (Radiochemical Centre) for 1 h at 37°C , were added in duplicate to 200 μl of effector cells in flat-bottomed microtitre plates (Linbro 96-TC), at effector:target (E:T) ratios of 80:1, 40:1, 20:1 and 10:1. Plates were incubated for 4 h at 37°C in a humidified atmosphere of 5% carbon dioxide in air. Macroassays were performed in triplicate in plastic tubes, where 4×10^4 target cells (labelled with $150 \mu\text{Ci } ^{51}\text{Cr}$ for 30 min at 37°C) were added to effector cells at E:T ratios of 100:1 and 50:1 in a total volume of 1 ml. The tubes were incubated as above for 12 h. Appropriate controls were included in both assays to determine total and spontaneous release of isotope. After incubation, 100 μl of supernatant was carefully collected from the microassay and 500 μl from the macroassay and counted in a γ -scintillation spectrometer (Packard). Results are expressed as the percentage of specific lysis calculated from the formula:

$$\% \text{ specific lysis} = 100 \times \frac{\text{c.p.m. in test} - \text{c.p.m. in spontaneous release}}{\text{total release c.p.m.} - \text{c.p.m. in spontaneous release}}$$

In addition, unidirectional MLRs were established using responder PBL from immune calves and stimulator cells from autologous infected cell lines which had been treated with $25 \mu\text{g ml}^{-1}$ mitomycin C (Sigma) for 30 min at 37°C . 4×10^7 PBL were added to 8×10^6 stimulator cells in 16 ml of culture medium (as above) supplemented with 10^{-4}M 2-mercaptoethanol, 100 IU ml^{-1} penicillin and $100 \mu\text{g ml}^{-1}$ streptomycin. Aliquots (2 ml) were dispensed into 24 well $\times$ 16 mm culture plates (Linbro) and incubated at 37°C in a humidified atmosphere of 5% CO_2 in air. After 5 d, the viable cells were recovered, washed and used in the microcytotoxicity assay, the figure depicts results from macrocytotoxicity test (a●) (50:1) direct (b●) and MLR-generated (c●) microcytotoxicity assays (40:1) using PBL effector cells against autologous (—) and allogeneic infected target cells (—○—), and effluent lymphatic lymphocytes (ELL) as effector cells against autologous infected targets (▲—▲).

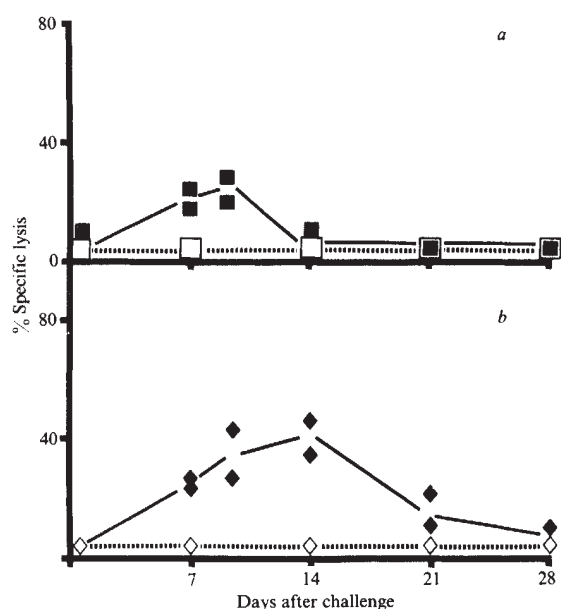


Fig. 3 Cytotoxic response of PBL from two immune calves after a second challenge with stablified infective sporozoites of *T. parva* which were inoculated subcutaneously 10 weeks after infection and treatment and 6 weeks after the first sporozoite challenge. Results are depicted direct (a■) and MLR-generated (b◆) microassays (40:1) using as targets autologous (—) and allogeneic (---) infected cell lines.

MLRs (data not shown). This result contrasted with the substantial autologous MLRs initiated by LNC during the second week of a lethal primary infection¹¹. When the challenge consisted of freshly isolated sporozoites adsorbed on autologous PBL, a transient parasitosis (<0.5%) was observed in efferent lymph from the draining lymph node on day 9. In these calves, the proportion of blast cells increased sharply from 5 to 30% on days 8–9 after challenge (Fig. 1). Cells from lymph nodes in which parasites were detected induced MLRs in normal cryopreserved autologous PBL. These results indicated that LNC are able to stimulate MLRs only when blasts, presumably infected cells, are present. Moreover, in immune animals challenged with sporozoites in the form of stablitate, the infection did not progress to a stage where macrophages could be detected by immunofluorescence or where infected cells could stimulate an autologous MLR.

Both challenge protocols with infective sporozoites elicited the development of cytotoxicity against autologous infected cells (Fig. 2a, b). Lytic activity was first detected in PBL 10–12 days after challenge, was maximal around day 14 when 20–50% of specific lysis was observed, and declined thereafter. Unlike in primary infections, no cytotoxicity was demonstrable against allogeneic cells or xenogeneic target cells at any stage. In one animal, lymphocytes draining from the regional lymph node exhibited cytotoxic activity against autologous infected cells from days 9–14, after the parasitosis had disappeared and before the detection of lytic activity in PBL (Fig. 2a, b). Calves re-challenged with a second inoculum of sporozoites as stablitate also showed cytotoxic responses in PBL which were again restricted to autologous target cells (Fig. 3a). These appeared at day 7, earlier than after the initial sporozoite challenge.

Similarly, strong cytotoxicity against autologous infected cells was found in three immune calves 6–14 days after challenge with 5×10^8 autologous infected cells (Fig. 4a, b). This lytic activity, which ranged between 20 and 60%, declined during the third week after challenge; once again, no activity was detected against allogeneic infected cells or xenogeneic tumour target cells. Moreover, the cytotoxicity was not due to any unique sensitivity to lysis of the allogeneic and autologous cell lines used: in experiments using crossed combinations of effector and target cells simultaneously, cytotoxicity was consistently observed against autologous but not allogeneic infected target cells

(Table 1). After challenge with autologous parasitized cells, leukocytes from two immune twin calves killed infected cells of their respective partners as efficiently as autologous cells (Table 1).

In other experiments, MLRs were performed using PBL from immune animals as responders and mitomycin-treated autologous infected cells as stimulators. When PBL were taken 21 days after challenge of immune animals with sporozoites and cultured for 5 days, a relatively low peak of cytotoxicity (~20%) was observed (Fig. 2c). When PBL were taken from immune cattle 7–24 days after challenge with autologous infected cells or with a second dose of stablitate, and cultured for 5 days, stronger cytotoxicity was detected (30–80%) (Figs 3b, 4c). In both cases, lysis was confined to autologous infected targets.

The present observations provide evidence that a population of specific cytotoxic cells is developed in cattle which have recovered from a primary infection with *T. parva*. These cells have the capacity to lyse autologous or cotwin but not allogeneic parasitized cells. The absence from peripheral blood of cytotoxic cells, and of precursors from which cytotoxic cells can be generated in MLR, at intervals other than those defined in this study, probably indicates that such cells reside in the solid lymphoid organs between encounters with the parasite. Following successive challenge with *T. parva*, specific cytotoxic cells enter central lymph and blood with a latent period progressively declining to a minimum of 6–7 days. Our findings demonstrate that cytotoxicity is effected directly by leukocytes collected from blood and lymph, discounting the necessity for an initial MLR to detect lytic activity⁷.

The genetic restriction of the cytotoxic response against *T. parva* in immune animals reported here contrasts with the lysis of allogeneic target cells reported by Pearson *et al.*⁷. The reasons for the discrepancy are not clear, but may reside in the multiple inoculations of autologous infected cells that those investigators

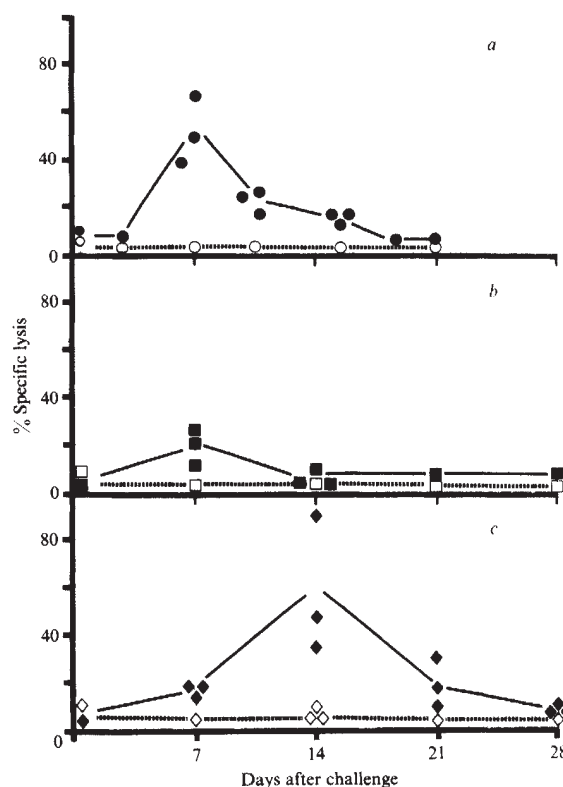


Fig. 4 Cytotoxic response of PBL from three immune calves at intervals after challenge with 5×10^8 autologous infected cells which were inoculated subcutaneously 10 weeks after immunization and 6 weeks after challenge with stablified *T. parva*. Results are depicted from direct macrocytotoxicity tests (a●) (50:1), direct (b■) and MLR-generated (c◆) microassays (40:1) using as targets autologous (—) and allogeneic infected cell lines (---).

used to obtain PBL which expressed cytotoxicity after an MLR. Repeated challenges of immune calves with parasitized autologous cells in our studies has only yielded cytotoxicity of autologous infected target cells. After challenge of each of 18 immune animals so far examined, killing of allogeneic targets has not been detected in direct assays against a range of parasitized allogeneic cell lines. Alternatively, the former workers used only a single immune animal and one allogeneic infected cell target, and it was not established that this particular combination was genetically unrelated; effector and target could have shared some of the relevant histocompatibility antigens.

The restriction of cytotoxicity by the cells of immune animals to autologous or chimaeric twin partner infected target cells is analogous to the restriction of T-lymphocyte-mediated cytotoxicity of murine virus-infected cells¹², or hapten-modified syngeneic targets¹³. This is the first time that such restriction has been observed for a protozoan parasite. By analogy, cytotoxic lymphocytes from immune cattle may recognize complexes of parasite-specific antigens and polymorphic host determinants (probably major histocompatibility antigens), on the surface of infected cells. Analysis of this interpretation will require more information about products of the major histocompatibility complex in cattle than is yet available. The kinetics of appearance and activity of specific cytotoxic cells suggest that these cells represent a major mechanism of recovery from infection. They are observed in animals that are being immunized by infection and treatment, and in immune animals challenged with sporozoites or macroschizont-infected cells. Recovered animals that are immune to the homologous isolate of the parasite are susceptible to some other isolates of *T. parva*¹⁴. If specific cytotoxicity is correlated with resistance, the effector cells would be expected to kill autologous cells infected with isolates of the parasite to which there is cross protection *in vivo*. Experiments are in progress to clarify this point, with the objective of developing a laboratory test to ascertain whether isolates or antigenic types of *T. parva* might be cross-protective *in vivo*, an important prerequisite for a vaccination programme.

One approach to immunoprophylaxis against *T. parva* envisages using infected cell lines. With respect to this approach, genetic restriction of the cellular response to *T. parva* implies that successful prophylaxis requires the parasite to become established in the lymphocytes of the host. The factors which govern the transfer of *T. parva* macroschizonts from allogeneic to host cells have not been elucidated. Although this exchange can occur, the process seems to be relatively inefficient, which may account for the large number (of the order of 10⁸) of allogeneic cells required to infect or immunize cattle. In contrast, infections can be initiated with as few as 100 autologous macroschizont-bearing cells¹⁵. Studies are in progress in this laboratory to characterize the antigenic heterogeneity of the various isolates of *T. parva* and to define the nature of the genetic restriction involved in the acceptability of parasitized allogeneic cells by the host for vaccination. Cellular immunity to parasites has previously been attributed to production by lymphocytes of factors that activate macrophages so that they limit the multiplication of parasites within them. We now describe a new mechanism: direct killing by lymphocytes of parasitized cells. This may well also apply in other intracellular parasite infections, along with the genetic restriction defined here.

We thank Dr G. Büscher for theilerial sporozoites, Messrs R. Nelson for *Theileria*-infected cell lines and Tom Tenywa for technical assistance, Pfizer Laboratories for the provision of Terramycin LA, Drs A. C. Allison, S. Black, W. I. Morrison and Max Murray for constructive discussions and support throughout these investigations, and Ms Leah Agina for typing the manuscript.

Received 28 August; accepted 17 December 1980.

1. Barnett, S. F. *Bull. Epizoot. Dis. Afr.* **5**, 543–557 (1957).
2. De Martini, J. C. & Moulton, J. F. *J. comp. Path.* **83**, 281–298 (1973).
3. Radley, D. E. *et al. Vet. Parasit.* **1**, 51–60 (1975).

4. Burridge, M. J., Morzaria, S. P., Cunningham, M. P. & Brown, C. G. D. *Parasitology* **64**, 511–515 (1972).
5. Wagner, G. G., Duffus, D. P. H. & Burridge, M. J. *Parasitology* **69**, 43–53 (1974).
6. Emery, D. L. *Res. vet. Sci.* (in the press).
7. Pearson, T. W., Dolan, T. T., Stagg, D. S. & Lundin, L. B. *Nature* **281**, 678–680 (1979).
8. Emery, D. L., Eugui, E. M., Nelson, R. & Tenywa, T. *Immunology* (in the press).
9. Brown, C. G. D., Stagg, D. A., Purnell, R. E., Kanhai, G. K. & Payne, R. C. *Nature* **245**, 101–103 (1973).
10. Cikes, M., Friberg, S. Jr & Klein, G. J. *Natn. Cancer Inst.* **50**, 347–350 (1973).
11. Emery, D. L. & Morrison, W. I. *Immunology* **40**, 229–237 (1980).
12. Doherty, P. C. & Zinkernagel, R. M. *J. exp. Med.* **141**, 502–507 (1975).
13. Shearer, G. M., Rehn, G. R. & Garbarino, C. A. *J. exp. Med.* **141**, 1348–1356 (1975).
14. Purnell, R. E. *Adv. Parasit.* **15**, 83–132 (1977).
15. Buscher, G., Morrison, W. I. & Nelson, R. (in preparation).
16. Eugui, E. M. & Allison, A. C. *Bull. Wld Hlth Org.* **57**, Suppl. 1, 231–238 (1979).

Normal levels of natural cytotoxic cells against solid tumours in NK-deficient beige mice

Osias Stutman & Michael J. Cuttito

Cellular Immunobiology Section, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

Natural cell-mediated cytotoxicity (NMC) capable of *in vitro* lysis of various lymphoid and non-lymphoid tumours has been described in mice and other species, including man¹. NCMC has been proposed as a first level of defence against tumour growth *in vivo*, one which does not need the priming of the conventional immunological response¹. The effector cells of NCMC seem to belong to a special category of lymphoid cells, being neither classical T or B cells nor macrophages^{1–5}; natural killer (NK) cells have been proposed as the prototype effector cell¹, although some heterogeneity among effector cells seems to exist, depending on the target cells used for testing^{4–10}. Two main subgroups of NCMC effector cells have been defined: NK cells directed against lymphoma targets^{2,3,6,7} and natural cytotoxic (NC) cells directed against solid non-lymphoid tumours^{4,5,8–10}. We describe here another distinction between the two systems: while NK activity is low in mice homozygous for the beige (*bg*) gene^{11–13} NC activity in spleen cell preparations from these animals is comparable with that observed in the appropriate controls (*bg*/+ and +/+ littermates). The *bg* syndrome of mice affects lysosome, melanosome and enzymatic functions and is a homologue of the Chediak-Higashi syndrome of man¹⁴. Defective NK activity in blood lymphocytes has also been reported in patients with Chediak-Higashi syndrome¹⁵. We also show that several mouse strains which have low NK activity, have normal or high levels of NC functions, expanding our previous observation that NC and NK cells are under distinct genetic control^{4,16}.

Table 1 summarizes the similarities and differences between NK and NC effector cells. As postulated earlier, it seems that NK and NC cells belong to a related, though distinct, family of effector cells, detectable in normal mice, which are capable of *in vitro* lysis of tumour cells without prior stimulation^{4,5,8}.

Table 2 shows our measurements of NK and NC activity of spleen cells against YAC-1 lymphoma and Meth-A sarcoma targets respectively, using C57BL/6J *bg*¹/*bg*¹ (line 4) and various controls (lines 1–3). Clearly NC activity is comparable in all groups studied, including the homozygous *bg* animals, while NK activity was detectable but significantly lower in the homozygous *bg* mice, in accordance with published reports^{11–13}. A small group of SB/Le mice which carry the *bg* gene derived from a radiation-induced mutation (the *bg*¹ being a spontaneous mutation)^{17,18} was also studied (lines 5 and 6). While NC levels between the homozygous *bg* and controls were comparable, NK activity was again markedly reduced in the homozygous *bg* (line 6). F₁ hybrids between C57BL/6J *bg*¹/*bg*¹ and SB/Le *sa* *bg*/*sa* *bg* also showed the same low NK reactivity (line 8), supporting the view that both *bg* mutations are acting on the same gene^{17,18}, and that the NK defect is not associated with the satin (*sa*) gene¹⁷.

In addition, Table 2 (lines 9–13) shows that other mouse strains such as A/J and some sublines³, SJL/J (S. Kaminsky and G. Cudkiewicz, personal communication) and I/St mice, which show low levels of NK activity when tested against YAC-1 targets, have normal levels of NC function against Meth-A sarcoma targets ('normal' means comparable with the syngeneic BALB/c strain⁴). Furthermore, I/St mice belong to the group of mouse strains with the highest NC reactivity against Meth-A targets^{4,8}.

The observation that transplanted tumours grow better and produce more metastases in homozygous *bg* mice^{19–21} supports the view of a possible *in vivo* surveillance role^{2,3} of the NK cells. On the other hand, one study, using a small group of homozygous *bg* and controls (8 mice per group), also showed that the incidence and time of appearance of sarcomas after administration of 3-methylcholanthrene were comparable²¹. This suggests that such 'normal' tumour incidence after chemical carcinogenesis in *bg* mice may be due to the adequate levels of NC cells present in such animals, as described here. Much larger groups of animals, as well as different carcinogen dosages have to be tested in homozygous *bg* mice and controls before this issue is settled. A similar speculation was made concerning the 'normal' incidence of spontaneous and induced tumours in athymic nude mice²², which have high levels of both NK^{2,3} and NC^{4,16} activity. The single published study on incidence of spontaneous tumours in *bg* mice¹⁷ showed that while the homozygous SB/Le *bg* animals had increased susceptibility to pneumonia, the incidence of lymphomas or other tumours was not increased. In fact, the incidences of lymphoproliferative lesions and lymphomas seemed to correlate more with the recessive gene *sa* than with *bg*¹⁷. Some cases of Chediak-Higashi syndrome in man develop a lymphoma-like tumour²³.

Similar comments can be made concerning the other mouse strains with low NK but normal NC activity described in Table 2 (lines 9–13). With the exception of SJL mice, which develop a high incidence of a reticulum cell lymphoid neoplasm²⁴, none of the other strains can be catalogued as having high tumour incidence^{24–27}. A/J mice show high incidence of lung adenomas at 18 months of age^{24,25} and mammary tumours in breeding females^{24,25} which are comparable with other mouse strains with normal NK activity. I/St mice in addition, are quite resistant to polycyclic hydrocarbon carcinogenesis^{26,27}. Again, as speculated, the 'normal' incidence of the common mouse tumours²⁴ in these strains with low NK activity (and also in the case of A/J mice with a deficient ability to generate tumoricidal macrophages²⁸) may be due to the normal NC activity observed in these animals. The low NK activity in the SJL may contribute to the high incidence of lymphoid tumours in such animals. On the other hand, additional factors may also be operative in SJL mice, as these animals were specifically bred for high incidence of the lymphoid tumours²⁴, and the incidence of spontaneous lymphomas is not clearly increased in mouse strains with low NK activity such as A sublines and 129/J (see ref. 3 for NK information and refs 24 and 27 for lymphoma incidence). Anti-cancer surveillance theories (whatever the mechanism for such surveillance) have predicted that an increase in tumour incidence would be the consequence of any spontaneous or induced interference with the surveillance device²⁹, but this was never clearly fulfilled for any surveillance model in animals or man^{27,30}.

It is tempting to speculate that NK–NC cells represent some form of alternative pathway of immunological surveillance ('alternative' meaning that it is not mediated by T cells as originally proposed in the theory²⁹, but that it is serving the same purpose of handling incipient tumours). Thus, animal models showing selective deficits of NK activity may become important for the determination of the *in vivo* role of such cells. However, using the same type of argument that we used in our criticism of the T-cell dependency of immunological surveillance^{22,27,30}, it may be that the effects detected in the NK–NC systems against tumours are only by-products of effector systems which serve other functions *in vivo*, perhaps related either to regulation of

Table 1 Comparison between NK and NC effector cells of natural cell-mediated cytotoxicity

Similarities	Differences
Pre-existing high levels in normal hosts, no need for apparent priming	Mouse strain distribution of activity
No evidence of immunological memory	Age of appearance
Levels of activity influenced by interferon	Affect of <i>in vivo</i> treatment with ⁸⁹ Sr, oestrogen, cyclophosphamide
Type of activity (that is, activity against otherwise negative targets) influenced by interferon and interferon inducers	Surface determinants (FcR, Thy-1, NK 1, Qa 2 and 5, Asialo GM1)
No H-2 restriction for cytotoxicity	
High reactivity dominant in F ₁ hybrids	
Present in athymic nude mice	
Tissue distribution	
No surface Ig and Ia negative	
Non-adherent	
Non-phagocytic	
Size	
<i>In vivo</i> effects of irradiation and steroids	
Blocking of cytotoxicity by monosaccharides	

NC^{4,5,8–10} and NK³ cells were compared using mostly NK data with YAC-1 targets. Some of the properties suggesting differences are not general: NK cells against EL4 lymphoma targets are not affected by *in vivo* treatment with ⁸⁹Sr (ref. 6), as are NC cells⁸; age of appearance of NC activity is usually earlier than NK, although that is dependent on the target cells used and in some cases activity is detected only at 3–4 weeks of age, as with NK⁹; with surface determinants, although NC cells are negative for those listed above^{5,8}, note that while some markers like Asialo GM1 seem to be present on all NK effector cells³⁴, other markers such as Thy-1, FcR or Qa 5 may be present only on fractions or subpopulations of NK cells^{37,38}. The surface phenotype of most NK cells is NK 1 (refs 7, 38), Qa 2 (refs. 38), Qa 5 (ref. 38) and Asialo GM1 (ref. 34) positive, while fractions of NK cells also express Thy-1 (refs 37, 38) and receptors for Fc (refs 2, 3). All these markers are absent from NC cells^{5,8,39}. Other markers present on NK cells not yet tested in NC include Ly 5, Ly 6 and receptors for *Helix pomatia*^{1–3,38}. Some of the similarities also deserve comment. For example, NC cells show a certain degree of adherence to nylon wool⁵ which was also observed with certain types of activated NK cells^{1–3}, although the bulk of NK activity is non-adherent to nylon. The blocking of NCMC by monosaccharides¹⁰ also shows differences: while NC cells show some selectivity to blocking by D-mannose (and galactose), NK cells show less selectivity and other physiological sugars can also block NCMC. For a review see ref. 39.

haematopoiesis, defence against virus or other infectious or parasitic agents^{1–3,8,14}.

Note three points concerning the present comparisons: (1) The NK defect observed in beige mice, although profound, is not total, and some degree of reactivity can be observed either in long-term assays (G. Cudkiewicz, personal communication) or after *in vivo* administration of interferon inducers^{19,31,32}. Thus, the present comparisons using a 4-h assay for NK but a 24-h assay for NC cells can be questioned. Kinetic studies of NC-mediated cytotoxicity have shown that most of the activity occurs within the first 12 h (refs 5, 8), there is no 'activation' of NC (or NK) cytotoxicity after 24 h in the assay and there is undetectable production of interferon during the 24-h assay (E. Lattime, W. Stewart III and O. S., unpublished results), suggesting that NC cells are not a special kind of *in vitro* activated NK cell. In addition, the differences in the expression of surface antigens between NC and NK cells also argues against such an interpretation, especially as activated NK cells also express such markers as Qa 5 and Asialo GM1 (refs 33, 34). (2) Although the beige mice seem to be defective solely in NK-mediated cytotoxicity^{11–13}, *in vitro* antitumour cytostasis and cytotoxicity by *in*

Table 2 NC and NK activity of spleen cells from C57BL/6J, homozygous *bg*^J, heterozygous *bg*^J and controls, and from other mouse strains with low NK activity (A/J and sublines, SJL/J and I/St)

Spleen cell donors	No. of mice tested	NC activity anti-Meth A			Nk activity anti-YAC-1		
		100	50	10	100	50	10
1. C57B/6J +/+	12 (12)	42 ± 5.1	33 ± 5.9	16 ± 6.3	37 ± 2.5	21 ± 1.3	14 ± 1.0
2. C57BL/6J <i>bg</i> ^J /+	3 (3)	42 (38–48)*	37 (33–41)	15 (15–16)	34 (31–38)	25 (20–33)	19 (12–28)
3. C57BL/6J <i>bg</i> ^J /+ and +/+ littermates	16 (12)	44 ± 4.8	34 ± 5.7	22 ± 5.5	30 ± 1.7	23 ± 1.5	13 ± 1.1
4. C57BL/6J <i>bg</i> ^J / <i>bg</i> ^J	16 (16)	41 ± 3.9	32 ± 4.4	20 ± 4.6	14 ± 1.1†	8 ± 0.9	4 ± 0.6
5. SB/Le +/+ and <i>bg</i> <i>sa</i> /++	5 (5)	53 ± 4.9	40 ± 4.7	27 ± 3.0	29 ± 1.9	19 ± 1.8	9 ± 1.1
6. SB/Le <i>bg</i> <i>sa</i> / <i>bg</i> <i>sa</i>	5 (5)	57 ± 4.7	33 ± 3.5	25 ± 2.8	4 ± 0.6	2 ± 0.6	0
7. (C57BL/6 × SB/Le)F ₁ +/+	10 (5)	55 ± 5.1	40 ± 4.1	29 ± 3.0	38 ± 2.3	25 ± 1.9	17 ± 1.3
8. (C57BL/6J × SB/Le)F ₁ <i>bg</i> ^J / <i>bg</i> <i>sa</i>	10 (9)	54 ± 4.3	42 ± 3.9	27 ± 2.9	7 ± 1.0	4 ± 0.6	2 ± 0.1
9. A/J	11 (6)	49 ± 2.5	26 ± 3.4	15 ± 1.4	11 ± 1.7	6 ± 1.1	5 ± 1.1
10. A/Boy	10 (5)	42 ± 3.3	26 ± 3.4	10 ± 1.2	10 ± 1.8	5 ± 0.7	2 ± 0.8
11. A.Tla ^b /Boy	10 (5)	39 ± 2.6	20 ± 4.3	17 ± 1.6	12 ± 3.4	9 ± 2.7	4 ± 1.5
12. SJL/J	6 (6)	39 ± 4.0	27 ± 2.6	10 ± 1.1	5 ± 0.9	3 ± 1.5	0
13. I/St	17 (10)	63 ± 3.0	39 ± 4.1	24 ± 2.7	5 ± 0.8	2 ± 0.7	0

All mice were 6–8 weeks old when used as spleen cell donors, and in most cases were simultaneously tested for NC and NK activity. The beige mice, controls and A/J mice were obtained from Jackson Laboratories; the SJL/J mice were a gift from Dr G. Cudkowicz; the A and its congenic pair A/Tla^b were obtained from Dr E. A. Boyse; the I/St mice were from our colonies⁴, as were the *bg* F₁ hybrids and controls. All values presented are mean ± S.E.M. of individual mice. Numbers in parentheses indicate number of mice tested simultaneously for NK activity. For details on the BALB/c sarcoma Meth A and the A/Sn lymphoma YAC-1 see refs 4, 5, 8, 10. NC activity was tested using ³H-proline-labelled adherent target cells, 100, 50 and 10 effector/target ratios and 24-h incubation as described in refs 4, 5, 8. NK activity was tested with ⁵¹Cr-labelled YAC cells in suspension, using the same effector/target ratios in 4-h release assays as described in ref. 10.

* As only three animals were tested, the results are presented as mean and range in parentheses.

† One animal had 32, 21 and 20% cytotoxicity at the three effector/target ratios tested. All the others had values of 15% or less at the 100:1 ratios.

in vivo activated macrophages is defective³⁵, while activity of *in vitro* activated macrophages and promonocytes of *bg* mice may be normal^{11–13,35}. Thus, the beige mice may have additional defects of cells, such as macrophages, putatively involved in defence against tumours, making *bg* an interesting model of NCMC deficiency, but certainly not the litmus test for definition of NK function *in vivo*. (3) The apparent tumour 'selectivity' of the NC-NK system for solid tumours versus lymphomas deserves comment. One proposed terminology even uses NK_L for the conventional antilymphoma activity and NK_S or NK_{solid} for a cell with similar characteristics as NC, although tested in a different assay⁷. 'Solid' is used throughout, in the traditional pathologist's sense, opposing it to the diffuse malignancies such as leukaemia, and not defining a particular state of matter. Such tumour-type preference of NC and NK may not be absolute. Metastatic spread to the lungs of any type of tumour may be controlled by NK-like cells^{19,36}, while tumour growth in other sites may be controlled by NC-like mechanisms, at least for certain solid tumours. The *in vivo* studies with beige mice show a deficient response to lymphomas²⁰ and melanomas¹⁹ (technically a solid tumour, although 'modified to be sensitive to NK cytotoxicity' by long-term *in vitro* passage¹⁹). Such deficiency is evidenced by an increase in local growth of the lymphomas²⁰ and an increase in spontaneous and artificial lung metastasis from solid tumours^{19,21}. On the other hand, local tumour growth of melanoma¹⁹ or the Lewis lung carcinoma³⁵, as well as the appearance of lymph node melanoma metastases¹⁹ were comparable in beige mice and controls. It is also tempting to speculate that the differences between tumour type and NC or NK cytotoxicity are due to the predominance on certain tumours (lymphomas compared with solid tumours) of determinants that are recognized by NC and NK cells respectively, perhaps related to either concentration or appropriate display of certain monosaccharides associated with membrane glycoproteins or glycolipids, as may be suggested by our studies of blocking of cytotoxicity, showing D-mannose preference for the NC system¹⁰.

Thus the beige mutation which affects NK cells has no effect on NC activity, a situation observed in other low-NK mouse strains such as A/J and sublines, SJL/J and I/St. Whether this situation explains the overall low tumour incidence of most of these strains, if indeed NCMC is considered as an antitumour surveillance device, needs further study.

This research was supported by grants CA-08748, CA-15988 and CA-17818 from the NIH and grant IM-188 from the American Cancer Society. We thank Mary L. Devitt for the animal care, Linda Stevenson for preparation of the manuscript, Drs Edmund Lattime, Stuart Macphail and Richard Miller for criticism and Dr G. Cudkowicz for providing unpublished results.

Received 21 October 1980; accepted 19 January 1981.

- Heberman, R. B. (ed.) *Natural Cell-mediated Immunity Against Tumors* (Academic, New York, 1980).
- Herberman, R. B. & Holden, H. T. *Adv. Cancer Res.* **27**, 305–377 (1978).
- Kiessling, R. & Wigzell, H. *Immun. Rev.* **44**, 165–208 (1979).
- Stutman, O., Paige, C. J. & Feo Figarella, E. *J. Immun.* **121**, 1819–1826 (1978).
- Paige, C. J., Feo Figarella, E., Cuttito, M. J., Cahán, A. & Stutman, O. *J. Immun.* **121**, 1827–1835 (1978).
- Kumar, V., Luevano, E. & Bennett, A. *J. exp. Med.* **150**, 531–547 (1979).
- Burton, R. C. in *Natural Cell-mediated Immunity Against Tumors*, 19–35 (Academic, New York, 1980).
- Stutman, O., Feo Figarella, E., Paige, C. J. & Lattime, E. C. in *Natural Cell-mediated Immunity Against Tumors*, 187–229 (Academic, New York, 1980).
- Stutman, O. in *Natural Cell-mediated Immunity Against Tumors*, 231–240 (Academic, New York, 1980).
- Stutman, O., Dien, P., Wisun, R. & Lattime, E. C. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2895–2898 (1980).
- Roder, J. & Duwe, A. *Nature* **278**, 451–453 (1979).
- Roder, J. *J. Immun.* **123**, 2168–2173 (1979).
- Roder, J., Lohmann-Matthes, M. L., Domzig, W. & Wigzell, H. *J. Immun.* **123**, 2174–2181 (1979).
- Padgett, G. A., Holland, J. M., Prieur, D. J., Davis, W. C. & Gorham, J. R. in *Animal Models for Biomedical Research* Vol. 3, 1–12 (National Academy of Sciences, Washington DC, 1970).
- Halliotis, T. et al. *J. exp. Med.* **151**, 1039–1048 (1980).
- Stutman, O. & Cuttito, M. J. in *Natural Cell-mediated Immunity Against Tumors*, 431–442 (Academic, New York, 1980).
- Lane, P. W. & Murphy, E. D. *Genetics* **72**, 451–460 (1972).
- Brandt, E. J. & Swank, R. T. *Am. J. Path.* **82**, 573–588 (1976).
- Talmadge, J. E., Meyers, K. M., Prieur, D. J. & Starkey, J. R. *Nature* **284**, 622–624 (1980).
- Karre, K., Klein, G. O., Kiessling, R., Klein, G. & Roder, J. C. *Nature* **284**, 624–626 (1980).
- Salomon, J. C., Creau-Goldberg, N. & Lynch, N. R. *Cancer Immun. Immunother.* **8**, 67–70 (1980).
- Stutman, O. in *The Nude Mouse in Experimental and Clinical Research*, 411–435 (Academic, New York, 1978).
- Dent, P. B., Fish, L. W., White, J. F. & Good, R. A. *Lab. Invest.* **15**, 1634–1642 (1966).
- Murphy, E. D. in *Biology of the Laboratory Mouse*, 521–562 (McGraw-Hill, New York, 1966).
- Heston, W. E. *J. natn. Cancer Inst.* **3**, 79–82 (1942).
- Strong, L. C. *Yale J. Biol. Med.* **25**, 34–43 (1952).
- Stutman, O. *Adv. Cancer Res.* **22**, 261–422 (1975).
- Boraschi, D. & Meltzer, M. E. *J. Immun.* **124**, 1050–1053 (1980).
- Burnet, F. M. *Prog. exp. Tumour. Res.* **13**, 1–27 (1970).
- Stutman, O. in *The Handbook of Cancer Immunology* Vol. 7, 1–25 (ed. Waters, H.) (Garland STPM, New York, 1980).
- Durdik, J. M., Beck, B. N. & Henney, C. S. in *Natural Cell-mediated Immunity Against Tumors*, 37–46 (Academic, New York, 1980).
- Brunda, M. J., Holden, H. T. & Heberman, R. B. in *Natural Cell-mediated Immunity Against Tumors*, 411–415 (Academic, New York, 1980).

33. Chun, M., Pasanen, V., Hammerling, U., Hammerling, G. F. & Hoffman, M. K. *J. exp. Med.* **150**, 426–431 (1979).
34. Young, W. W., Hakomori, S. I., Durkik, J. M. & Henney, C. S. *J. Immun.* **124**, 199–201 (1980).
35. Mahoney, K. H., Morse, S. S. & Morahan, P. S. *Cancer Res.* **40**, 3934–3939 (1980).
36. Riccardi, C., Santoni, A., Marlozzari, T. & Herberman, R. B. in *Natural Cell-mediated Cytotoxicity Against Tumors*, 1121–1139 (Academic, New York, 1980).
37. Mattes, M. J., Sharrow, S. O., Herberman, R. B. & Holden, H. T. *J. Immun.* **123**, 2851–2890 (1979).
38. Koo, G. C., Jacobson, J. B., Hammerling, G. J. & Hammerling, U. *J. Immun.* **125**, 1003–1006 (1980).
39. Stutman, O., Lattime, E. C. & Feo Figarella, E. *Fedn Proc.* (in the press).

Idiotypic regulation by isologous monoclonal anti-idiotope antibodies

M. Reth, G. Kelsoe & K. Rajewsky

Institute for Genetics, University of Cologne, Weyertal 121, D-5000 Köln 41, FRG

The network hypothesis¹ proposes that internal antibody interactions are crucial in the regulation of the immune system. Thus, injection of minute amounts of anti-idiotypic antibodies (antibodies specific for a particular set of antigenic determinants (idiotopes) associated with the variable region of a particular antibody or set of antibodies) can enhance or suppress the production of the corresponding set of idiotopes (the idiotype) in subsequent antibody responses of the animal^{2–5}. We have recently shown that idiotope expression can be enhanced and suppressed by monoclonal anti-idiotope antibodies of the IgG1 class in concentrations approaching those of natural antibodies⁶. We now demonstrate that these regulatory effects are not due to differences in the constant parts of the injected antibody and antibodies of the IgG1 class in the recipient. Also, an IgG2a anti-idiotope antibody raised in C57BL/6 mice suppresses the expression of the corresponding idiotope in animals of the same strain when minute amounts are injected 2–8 weeks before the test immunization with antigen. These results indicate that antibody-mediated network interactions are physiologically important and suggest that the class of isologous anti-idiotypic antibodies may influence their regulatory properties. Finally, the enhancement and suppression of idiotope expression by IgG1 anti-idiotopes and the suppression by IgG2a is almost exclusively at the level of IgG but not IgM antibodies.

We developed previously, by the Köhler–Milstein cell hybridization technique⁷, a series of monoclonal antibodies of mouse origin each of which recognizes a particular idiotope characteristic for a subset of anti-NP (NP = 4-hydroxy-3-nitrophenyl-acetyl) antibodies of C57BL/6 mice⁸. When IgG1 antibodies of this type were injected into C57BL/6 mice and the animals immunized with an NP-carrier conjugate 1–8 weeks later, the expression of the corresponding idiotope in the anti-NP antibody population was either enhanced or suppressed, depending on the dose of anti-idiotopic antibody: nanogramme amounts led to enhancement, microgramme amounts to suppression (ref. 6 and unpublished observations).

The antibodies used in these experiments originated from CBA mice and thus differed in allotype from IgG1 antibodies of the C57BL/6 strain. The experiment was therefore repeated in (C57BL/6 × CBA)F₁ animals for which the anti-idiotope antibody would be 'self', at least with respect to its constant part. Groups of five animals were injected with 10 ng or 10 µg antibody Ac38 and immunized with NP₈-CG (CG = chicken γ-globulin) 4 or 6 weeks later. The experimental conditions were exactly as described in our previous paper⁶, except that the anti-idiotope antibodies were diluted in phosphate-buffered saline containing 5 µg mouse serum albumin per ml. The animals were bled 12 days after immunization with antigen and the sera were titrated for λ₁-bearing anti-NP antibodies and the

Ac38 idiotope (the idiotope recognized by antibody Ac38). This idiotope is expressed within the λ-bearing antibody population which represents a major portion of the total primary anti-NP response in mice expressing the *Igh^b* haplotype^{9,10}. The results of the experiment are depicted in Fig. 1.

It is clear that 10 ng of antibody Ac38 enhance and 10 µg of the antibody suppress the expression of the Ac38 idiotope in the primary anti-NP response. There is no effect on the total amount of λ₁-bearing anti-NP antibodies. These results are strictly analogous to those previously obtained in C57BL/6 mice⁶ and thus demonstrate that allotype disparity is not responsible for the regulatory effect of these anti-idiotope antibodies.

Sera from one of the two experiments depicted in Fig. 1 were analysed for the class distribution of idiotope-bearing and total anti-NP antibodies. The sera from the control groups were not available for analysis, as they had been used up in absorption experiments in which it was shown that elimination of NP-binding molecules from the sera also led to the elimination of more than 90% of idiotope-bearing molecules. We therefore included in the analysis sera from our previous experiments⁶, in which antibody Ac38 had been injected into C57BL/6 mice instead of (CBA × C57BL/6)F₁ animals. The results are shown in Fig. 2.

The sera from the F₁ experiment were titrated for IgM, IgG1, IgG2a and IgG2b (Fig. 2a). IgG2a anti-NP antibodies were not detectable and IgG1 was the predominant class. The level of antibodies bearing the Ac38 idiotope was high in IgG in the 10-ng group and reduced to background in the 10-µg group.

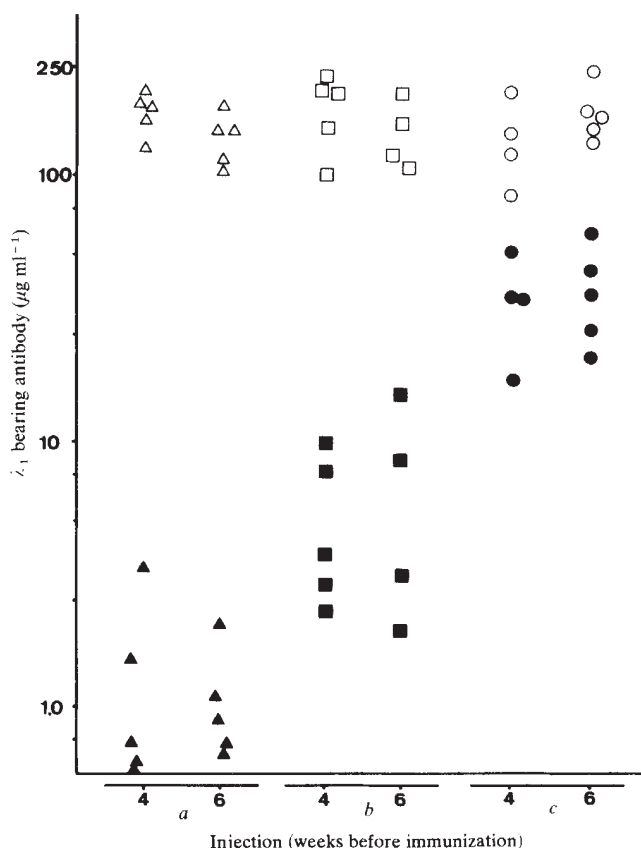


Fig. 1 Serum antibody levels from (CBA × C57BL/6)F₁ mice injected with: a, 10 µg Ac38 (△, ▲); b, diluent (□, ■); and c, 10 ng Ac38 (○, ●) 4 or 6 weeks before immunization with 100 µg NP₈-CG. Mice were bled 12 days after immunization and individual serum samples collected for analysis. Individual values from the results of a solid-phase radioimmunoassay are plotted as µg ml⁻¹ of λ₁-bearing antibody⁶. Open symbols represent the total amount of NP-specific antibody, closed symbols represent that fraction of antibody bearing the Ac38 idiotope.

In contrast, at the level of IgM, both groups exhibited similar levels of Ac38-positive anti-NP antibodies. The notion that the regulation of idiotope expression by antibody Ac38 operates mainly at the level of the IgG response is further supported by the data of the experiment with C57BL/6 mice (Fig. 2b). In this case the sera were analysed for idiotope expression in IgM and IgG1 as the major IgG component. Obviously enhancement of Ac38 expression appears mainly in IgG1 and not at all in IgM; suppression is complete in IgG1 and only partial in IgM.

One might still argue that the experiments documented in Figs 1 and 2 do not reflect a physiological process in the immune system because antibody Ac38 was not induced in the strain expressing the Ac38 idiotope (C57BL/6), although this argument does not seem to make much sense in view of the extreme diversity of the antibody system. Nevertheless, we have started to produce, in the same idiotypic system, monoclonal anti-idiotope antibodies of C57BL/6 origin. For this purpose C57BL/6 mice were sensitized with a conjugate of keyhole limpet haemocyanin and the anti-NP antibody B1-8, a monoclonal IgM of C57BL/6 origin—the same protein against which our previous anti-idiotope antibodies had been raised⁸. The animals were boosted with the same conjugate 2 weeks later and their spleen cells fused with the myeloma line X63.Ag8-6.5.3 (ref. 11) 3 days after boosting. All technical details of immunizations, the cell fusion technique⁷, the detection, cloning and propagation of cells secreting monoclonal anti-idiotope antibodies, the characterization of these antibodies and their purification have been described previously⁸. A cell line designated A6.24 was isolated, cloned and propagated in (C57BL/6 × BALB/c)F₁ animals; it secreted large amounts of a

monoclonal anti-B1-8 antibody. The antibody was purified from ascites fluid by ammonium sulphate precipitation and subsequent chromatography on DEAE cellulose (DE52, Whatman) as described previously⁸. The purified material showed a single band in microzone electrophoresis, and on SDS-polyacrylamide electrophoresis¹² in reducing conditions, two main bands appeared in the position of immunoglobulin L and H chains. It typed as an IgG2a in an enzyme linked immunosorbent assay¹¹. In radioactive binding assays antibody A6.24 reacted with the B1-8 protein and its Fab and Fv fragment⁸ and a fraction of C57BL/6 anti-NP serum antibodies (see below), but not with any of a series of myeloma proteins or, at a detectable level, with immunoglobulins from normal sera. The reaction of antibody A6.24 with B1-8 could be specifically blocked by an excess of NIP-BSA (4-hydroxy-5-iodo-3-nitrophenyl-bovine serum albumin) but not of NIP-caproic acid. Antibody A6.24 thus behaves as a typical anti-idiotope antibody detecting a group 2 idiotope in our previous nomenclature⁸. This idiotope is distinct from the idiotope defined by antibody Ac38, as antibodies A6.24 and Ac38 exhibit different patterns of reactivity with a series of NP-binding monoclonal antibodies (data not shown).

Antibody A6.24 is also distinguished from the previously used antibody Ac38 by its origin (C57BL/6) and by its class: it types as an IgG2a whereas Ac38 is an IgG1. How does injection of antibody A6.24 into C57BL/6 mice influence the expression of the corresponding idiotope (which we call idiotope A6.24) in the anti-NP response? Groups of age- and sex-matched C57BL/6 mice received 33 ng or 10 µg A6.24 intraperitoneally. Controls were injected with diluent alone. 2–8 weeks later the animals were immunized with NP₈-CG and bled 12 days later. The protocol of the experiment thus followed exactly that of our previous, analogous experiments (see above and ref. 6). The sera were titrated for total λ_1 -bearing anti-NP antibodies and for idiotope A6.24. The results of these titrations appear in Fig. 3.

It can be seen that, as in the experiments with Ac38, 10 µg A6.24 profoundly suppress the expression of the corresponding idiotope, leaving the total anti-NP response untouched. Suppression is almost complete throughout the entire experiment. However, at the 33-ng dose which would have led to maximal enhancement in the case of Ac38⁶, antibody A6.24 again produces suppression, although less pronounced than at the 10-µg dose. As previously⁶, there is an indication of oscillations of idiotope expression. The class distribution of the anti-NP antibodies in the sera from various groups of this experiment was also determined (data not shown). It was essentially the same as that in the experiment depicted in Fig. 2. Only idiotope-bearing IgG but not IgM antibodies were significantly suppressed.

Our data show that striking modulation of idiotope expression in an immune response can be achieved by the prior administration of an isologous, monoclonal anti-idiotope antibody. Thus, the capacity to recognize, respond against and regulate the expression of a given idiotypic determinant exists within a single mouse strain. This strongly suggests that antibody-mediated network regulation is an important physiological process as discussed in detail elsewhere⁶. Our data also suggest that the effector function of isologous anti-idiotope antibodies in network regulation may be influenced by the heavy-chain constant region. Antibody Ac38, an IgG1, enhances and suppresses, depending on the dose. The same has been found for another IgG1 anti-idiotope, Ac146 (ref. 13). Antibody A6.24, an IgG2a, may only suppress. This is reminiscent of the previous finding that guinea pig IgG1 anti-idiotypic antibodies enhance idiotope expression in the mouse, whereas IgG2 has suppressive effects^{2,3}. At least in terms of their complement-fixing properties, IgG1 and IgG2 in mice and guinea pigs are similar^{14,15}. Clearly, however, our experiments can not formally assign distinct functions in network control to the constant regions of γ_1 and γ_2a heavy chains. It is known, for example, that different allelic forms of the γ_2a chain may differ drastically from each other in primary structure (P. Schreier and

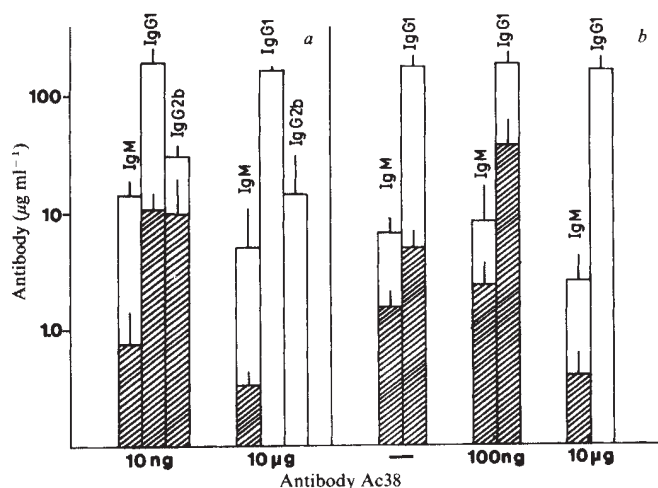


Fig. 2 Immunoglobulin class distribution of total and idiotope Ac38-bearing anti-NP antibodies in the sera of groups of mice injected with different doses of anti-idiotope antibody (Ac38) and subsequently immunized with 100 µg NP₈-CG. Each column represents the geometric mean ($\pm$ s.d.) of the amounts of anti-NP antibody of a particular immunoglobulin class expressed within a group of mice. That fraction of antibodies bearing the Ac38 idiotope is given by the shaded portion of each column. *a*, (CBA × C57BL/6)F₁ mice injected with 10 ng or 10 µg of Ac38 6 weeks before immunization with NP-CG. *b*, C57BL/6 mice injected with diluent, 100 ng Ac38, or 10 µg Ac38. Based on earlier results⁶, mice injected with 10 µg or 100 ng Ac38 were immunized with NP-CG 4 or 8 weeks later, respectively. Antibody concentrations were determined in a solid-phase radioimmunoassay⁶. To determine total anti-NP antibody, hapten-binding immunoglobulin was detected with ¹²⁵I-labelled, immunoglobulin class-specific antibodies (described in ref. 16). Those antibodies bearing the Ac38 idiotope were detected in a similar manner with the exception that the solid-phase coat was the F(ab)₂ fragment of antibody Ac38 rather than a hapten-protein conjugate. Absorption experiments showed that virtually all (>90%) idiotope Ac38-positive antibodies bound the NP hapten.

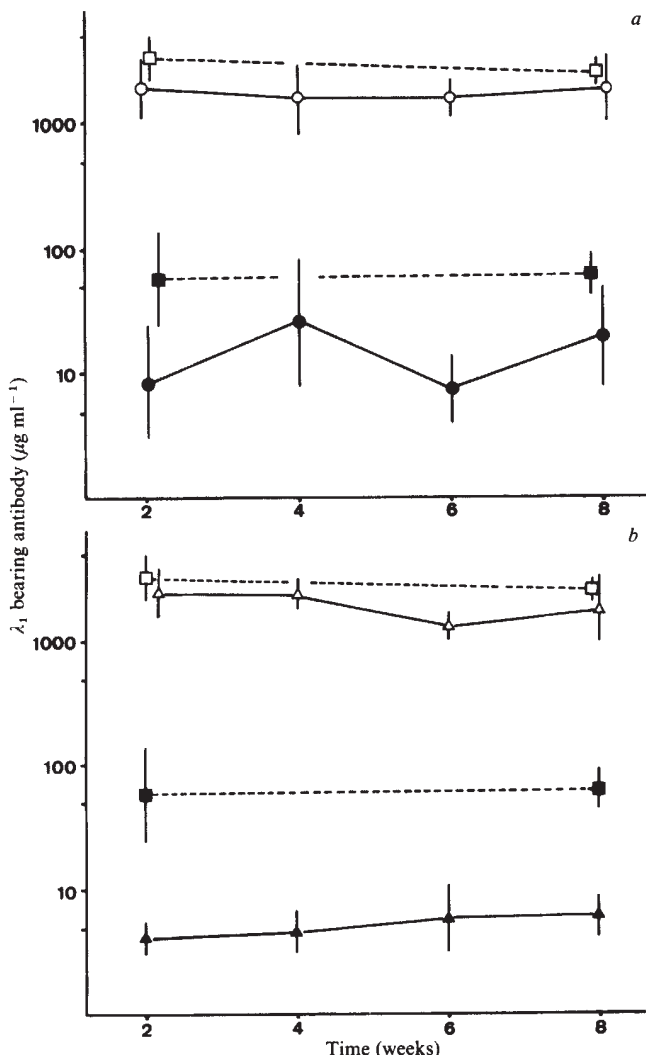


Fig. 3 Idiotope suppression induced by monoclonal anti-idiotope antibody A6.24. Groups of six mice were injected intraperitoneally with *a*, 33 ng (○, ●; unbroken lines) or *b*, 10 µg (△, ▲; unbroken lines) of antibody A6.24 and immunized 2–8 weeks later with NP-CG. Control mice (□, ■; broken lines) received diluent instead of anti-idiotope antibody. 12 days after the administration of antigen the animals were bled and the sera of individual mice analysed in a solid-phase radioimmunoassay⁶ for total, λ₁-bearing anti-NP antibodies (open symbols) and λ₁-bearing anti-NP antibodies expressing the A6.24 idiotope. The geometric means (with s.d.) of the antibody concentrations in the sera are plotted against the time between injection of anti-idiotope antibody and antigen.

A. Bothwell, personal communication). A definitive assignment will require the testing of a number of isologous anti-idiotope antibodies of various isotypes and (more importantly) the isolation of pairs of anti-idiotope antibodies with identical variable regions but belonging to different immunoglobulin classes, for example class-switch variants. A technique for the isolation of such pairs of antibodies is available¹⁶. Finally, it seems that the target of regulation by our anti-idiotope antibodies is mainly the IgG and not the IgM antibody response. It will be important to investigate the basis of this finding as regulatory studies with conventional isologous anti-idiotypic antisera suggest that in certain conditions, IgM and IgE responses can also be affected by idiotypic control^{4,17}.

This investigation was supported by the Deutsche Forschungsgemeinschaft through SFB 74. G.K. is a recipient of a Fellowship from the US NSF. We thank Ms Gertrud von Hesberg for technical assistance.

Received 26 September 1980; accepted 20 January 1981.

1. Jerne, N. K. *Annls. Immun. Inst. Pasteur, Paris* **125C** 373 (1974).
2. Eichmann, K. *Eur. J. Immun.* **4**, 296 (1974).
3. Eichmann, K. & Rajewsky, K. *Eur. J. Immun.* **5**, 661 (1975).
4. Köhler, H. *Transpl. Rev.* **27**, 24 (1975).
5. Julius, M. H., Augustin, A. A. & Cosenza, H. in *Regulatory Genetics of the Immune System* (eds Sercarz, E. E., Herzenberg, L. A. & Fox, C. F.) Vol. 6, 179 (Academic, New York, 1977).
6. Kelsoe, G., Reth, M. & Rajewsky, K. *Immun. Rev.* **52**, 75 (1980).
7. Köhler, G. & Milstein, C. *Nature* **256**, 495 (1975).
8. Reth, M., Imanishi-Kari, T. & Rajewsky, K. *Eur. J. Immun.* **9**, 1004 (1979).
9. Imanishi-Kari, T., Rajnavölgyi, E., Takemori, T., Jack, R. S. & Rajewsky, K. *Eur. J. Immun.* **9**, 324 (1979).
10. Karjalainen, K. *Eur. J. Immun.* **10**, 132 (1980).
11. Kearney, J. F., Radbruch, A., Liesegang, B. & Rajewsky, K. *J. Immun.* **123**, 1548 (1979).
12. Laemmli, K. K. *Nature* **227**, 680 (1970).
13. Kelsoe, G., Reth, M. & Rajewsky, K. *Eur. J. Immun.* (in the press).
14. Klaus, G. G. B., Pepys, M. B., Katajima, K. & Askonas, B. A. *Immunology* **38**, 687 (1979).
15. Bloch, K. J., Kourilsky, F. M., Ovary, Z. & Benacerraf, B. *J. exp. Med.* **117**, 965 (1963).
16. Radbruch, A., Liesegang, B. & Rajewsky, K. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2909 (1980).
17. Blaser, K., Nakagawa, T. & de Weck, A. *J. Immun.* **125**, 24 (1980).

Control of enzyme secretion by non-cholinergic, non-adrenergic nerves in guinea pig pancreas

G. T. Pearson, J. S. Davison, R. C. Collins & O. H. Petersen*

Department of Physiology, University of Dundee, Dundee DD1 4HN, UK

Depolarization of pancreatic cells by exposure to high potassium solutions is associated with release of amylase. In the guinea pig, but not the mouse or cat, this Ca-dependent amylase secretion is resistant to atropine blockade^{1–3}, thus Scheele and Haymovits^{3,4} concluded that the enzyme secretion evoked by K depolarization does not involve release of transmitter from intrapancreatic nerves but is a consequence of Ca uptake into acinar cells mediated by the membrane depolarization. This hypothesis is inconsistent with current concepts of stimulus-secretion coupling in electrically non-excitable cells⁵. The observation of Scheele and Haymovits could, however, also be explained by the release of a non-cholinergic, secretomotor transmitter as a consequence of the depolarization of intrapancreatic nerves. By adapting the technique of electrical field stimulation of isolated pancreatic segments^{6,7} to our studies of amylase secretion, we have now been able to demonstrate both cholinergic and non-cholinergic, non-adrenergic secretomotor nerves in the guinea pig pancreas. Excitation of the non-cholinergic nerves stimulates amylase secretion by a different intracellular coupling mechanism from that activated by cholinergic nerves or by peptides belonging to the cholecystokinin, gastrin or bombesin families⁸.

All experiments were done on isolated segments of guinea pig pancreas superfused with oxygenated physiological salt solutions at 37 °C (ref. 7). Amylase secretion from the pancreas was measured using an on-line automated fluorometric method⁹, intracellular recordings of acinar transmembrane potentials were obtained using glass microelectrodes⁷, and Ca efflux was monitored by measuring the release of ⁴⁵Ca from prelabelled tissue¹⁰.

Electrical field stimulation (FS) has recently been shown to be an effective method of exciting intrinsic nerves in the pancreas^{6,7}. Using Perspex flow cells incorporating silver electrodes, we have studied the effects of FS on membrane potentials, amylase secretion and ⁴⁵Ca²⁺ efflux. In preliminary experiments we confirmed the results of Scheele and Haymovits^{3,4} that an elevated extracellular K concentration (75 mM) stimulated amylase secretion even in the presence of atropine (10⁻⁵ M). This protocol, however, is unhelpful as it is impossible to distinguish between indirect effects mediated by depolarizing nerve endings and direct effects due to depolariza-

* To whom correspondence should be addressed.

tion of the acinar cells. The nerve-blocking drug tetrodotoxin (TTX) (10^{-6} M) did not abolish the effects, as also noted by Scheele and Haymovits⁴, but K depolarization of nerve endings might of course evoke transmitter release even in the absence of action potential formation. We decided, therefore, to investigate more directly the possible existence of non-cholinergic nerves.

Figure 1a shows FS-evoked amylase secretion from a tissue superfused with normal physiological solution. Addition of the muscarinic antagonist atropine (10^{-5} M), in a concentration sufficient to abolish maximum response to exogenous acetylcholine (ACh) stimulation (10^{-4} M), reduced, but did not abolish, the FS response. Such an atropine-resistant amylase

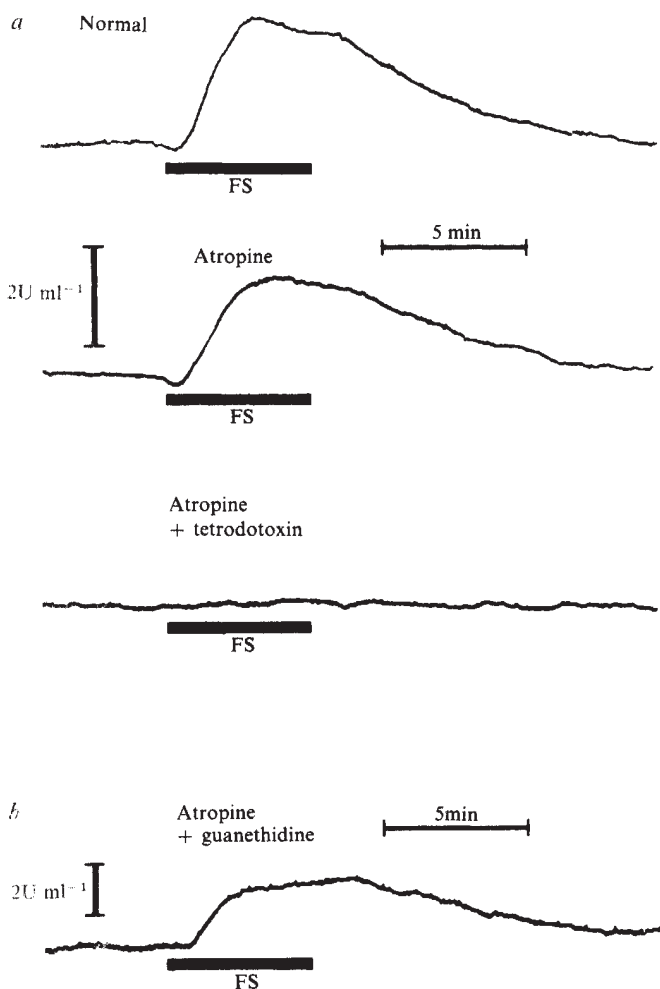


Fig. 1 Effects of electrical field stimulation (FS) on amylase release from superfused guinea pig pancreatic fragments. The figure shows two separate experiments (*a* and *b*) in which amylase in the effluent from the tissue chamber (1 ml vol, flow rate 1 ml min⁻¹) was continuously assayed by an automated fluorometric, on-line method⁹. FS, for the duration indicated, was achieved by connecting silver electrodes in the chamber to an isolated stimulator. Stimulation parameters were the same throughout: 20 Hz frequency, 50 V amplitude, 1 ms pulse width for 5 min. (FS using these parameters, even for periods up to 20 min, did not cause any change in the pH of the Krebs solution flowing through the tissue chamber.) The three panels in *a* are consecutive traces from the same experiment showing the effects of FS on the amylase concentration in the tissue chamber effluent during exposure of the tissue to a normal physiological salt solution, in the presence of atropine (10^{-5} M) and in the presence of both atropine (10^{-5} M) and TTX (10^{-6} M); *b* shows the FS response in the presence of both atropine (10^{-5} M) and guanethidine (10^{-5} M). All drugs applied were present in the superfusing solution for a minimum period of 20 min before FS. Note the different calibrations (2 units of amylase ml⁻¹) in *a* and *b*.

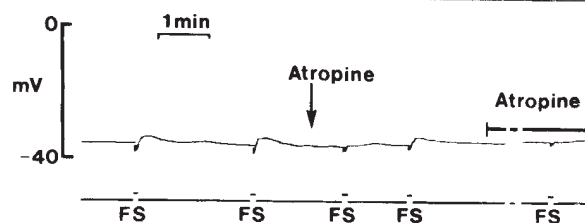


Fig. 2 Recording of guinea pig acinar cell membrane potential. An isolated segment of pancreas was secured to a Perspex platform and superfused with oxygenated physiological salt solution at 37 °C. A single microelectrode filled with 3 M KCl was inserted in an acinar cell to record a resting membrane potential of -35 mV. FS, as described in detail elsewhere⁷, was achieved using a pair of silver wire electrodes placed in contact with the tissue on the Perspex platform. These were connected to an isolated stimulator and the stimulation parameters throughout the experiment were 40 Hz, 30 V, 1 ms for 5 s. The effect of FS on the cell membrane potential is seen both in the absence and presence of atropine (10^{-5} M). At the arrow the solution flowing into the bath contained 10^{-5} M atropine for only ~10 s, thus the bath concentration of atropine is unlikely to have attained the 10^{-5} M concentration. Nevertheless, as seen in the trace, the FS effect was abolished. The response, however, rapidly returned and subsequent, sustained exposure to 10^{-5} M atropine (bar) again abolished the effect of FS. The break in the recording represents a 2-min interval.

secretion has been observed in 16 separate guinea pig preparations. Subsequent addition of TTX in a concentration that does not affect ACh or caerulein responses in the pancreas⁷ completely abolished the effect of FS. Abolition of the atropine-resistant FS response by TTX was observed in all six preparations tested, indicating that FS causes amylase release by initiating action potentials in intrapancreatic nerves. The adrenergic neurone-blocking drug, guanethidine (10^{-5} M), had no effect on the nerve-mediated amylase secretion in the five preparations tested (Fig. 1*b*), suggesting that the non-cholinergic response is not mediated by stimulation of adrenergic nerves. Furthermore, we have been unable to demonstrate any effect of noradrenaline (6×10^{-5} M) on amylase secretion. Although most experiments measuring nerve-mediated amylase secretion have been done using FS at 20 Hz frequency, effects have been observed at frequencies as low as 5 Hz. The mean value ($\pm$ s.e.m.) for FS-evoked amylase release at 20 Hz in the presence of atropine was $0.45(\pm 0.06)$ U per ml per 100 mg ($n=10$), corresponding to 20% of the mean value for the maximal secretion evoked by exogenous ACh (10^{-4} M). We conclude that the amylase secretion evoked by FS in the presence of atropine (Fig. 1) is mediated by excitation of intrapancreatic nerves, releasing a transmitter substance which activates the pancreatic acinar cells. The transmitter is not ACh and there is no indication that it is a catecholamine.

To test the functional innervation of individual acinar units⁷, membrane potential changes in response to FS were studied. Figure 2 shows examples of membrane depolarizations evoked by FS (40 Hz). Of 84 units studied, only 23 responded to FS, in marked contrast to the 100% effectiveness in the mouse⁷ and rat pancreas¹¹. The FS-evoked depolarizations, shown in Fig. 2, are always abolished by atropine (five preparations). It seems, therefore, that the transmitter released by excitation of the non-cholinergic, non-adrenergic nerves is unable to evoke membrane depolarization in spite of being a potent secretagogue.

The pancreatic secretagogues evoking enzyme secretion mediated by changes in cellular calcium metabolism (for example, ACh, cholecystokinin and bombesin) also cause acinar cell depolarization, while the secretagogue secretin (or vasoactive intestinal polypeptide, VIP) has no effect on ⁴⁵Ca fluxes or membrane electrophysiological parameters^{8,12,13}. We therefore tested the ability of FS to alter calcium metabolism in the guinea pig pancreas.

In the absence of atropine, FS caused a clear increase in fractional ^{45}Ca efflux from preloaded pancreatic segments in two separate experiments (Fig. 3). In the presence of atropine, however, the same type of FS (20 Hz, 40 V, 1 ms) failed to evoke any measurable effect on ^{45}Ca efflux in the four different preparations tested, although all of these responded with a marked increase in ^{45}Ca outflux when subsequently challenged with caerulein (1.5×10^{-9} M, Fig. 3).

The nerve-mediated secretion obtained in the presence of atropine (Fig. 1) is clearly not associated with membrane depolarization (Fig. 2) or with an increase in ^{45}Ca efflux (Fig. 3). The as yet unidentified transmitter substance is therefore unlikely to be a peptide belonging to the cholecystokinin-gastrin-caerulein, bombesin or substance P families because all secretagogues studied from these groups characteristically evoke marked Ca outflux and membrane electrical changes^{8,12,13}.

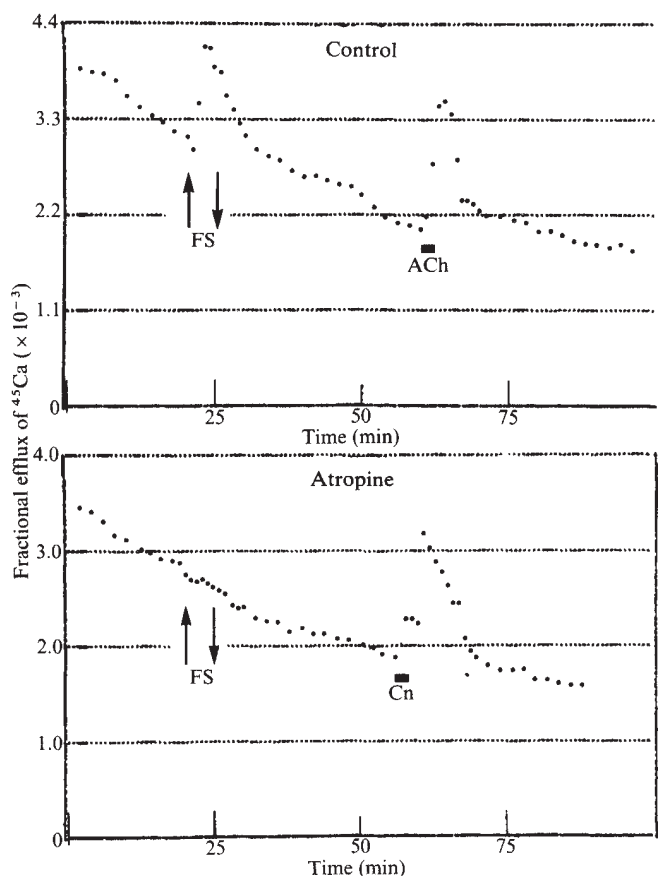


Fig. 3 The effect of FS and the secretagogues caerulein (Cn) and acetylcholine (ACh) on the fractional efflux of ^{45}Ca from preloaded superfused fragments of guinea pig pancreas. The pancreatic segments were loaded for 60 min in 2 ml of physiological salt solution containing $10 \mu\text{Ci } ^{45}\text{Ca}$ and maintained at 37°C . The procedure was similar to that described elsewhere¹⁰. After loading, the segments were transferred to a flow cell. In each experiment the tissue was superfused for 150 min before collection of effluent samples, which were subsequently analysed in a scintillation counter. Values for concentration of radioactivity in the samples taken at different time points together with the value for the radioactivity remaining in the tissue at the end of the experiment were processed by computer (MINC-11, Digital). Values were obtained for fractional efflux ($\Delta x / \Delta t$, min^{-1} , where x represents c.p.m. ^{45}Ca released in the time interval Δt and x , the tissue ^{45}Ca content at the midpoint of interval Δt) as a function of time. FS was carried out in exactly the same way as described for the amylase studies. Stimulation parameters were similar in each experiment: 20 Hz, 40 V, 1 ms for 5 min. Cn (1.5×10^{-9} M) and ACh (10^{-5} M) were added to the superfusing solution for the periods indicated by the bars. Atropine (10^{-5} M) was present throughout the experiment shown in the lower panel. Upper panel, normal preparation.

The present study has demonstrated, for the first time, the existence of non-cholinergic, non-adrenergic nerves controlling pancreatic acinar enzyme secretion. Its importance lies in the widespread distribution of non-cholinergic, non-adrenergic nerves throughout the peripheral autonomic nervous system^{14,15}. The transmitter released from the non-cholinergic, non-adrenergic nerves in our experiments might be VIP or a related peptide, as VIP-immunoreactive nerves have been demonstrated in the pancreas^{16,17}. VIP is known to stimulate amylase secretion by a mechanism not involving changes in cellular Ca metabolism^{8,13}.

Our results also resolve an important controversy in the discussion of pancreatic stimulus-secretion coupling⁸ because the Ca-dependent depolarization-evoked enzyme secretion observed in the presence of atropine^{3,4} can now most conveniently be explained as being due to Ca-dependent secretion of a neurotransmitter that interestingly does not seem to activate secretion through changes in cellular calcium metabolism.

This work was supported by grants from the MRC to O.H.P. We thank Drs I. Findlay and A. R. Chipperfield for advice and Dr M. S. Daoud and Miss K. Watson for technical assistance.

Received 5 November 1980; accepted 2 February 1981.

1. Argent, B. E., Case, R. M. & Scratcherd, T. *J. Physiol., Lond.* **216**, 611-624 (1971).
2. Poulsen, J. H. & Williams, J. A. *Pflügers Arch. ges. Physiol.* **370**, 173-177 (1977).
3. Scheele, G. & Haymovits, A. *Ann. N.Y. Acad. Sci.* **307**, 648-652 (1978).
4. Scheele, G. & Haymovits, A. *J. Biol. Chem.* **255**, 4918-4927 (1980).
5. Petersen, O. H. *Physiol. Rev.* **56**, 535-577 (1976).
6. Nishiyama, A., Katoh, K., Saitoh, S. & Wakui, M. *Membr. Biochem.* **3**, 49-66 (1980).
7. Davison, J. S., Pearson, G. T. & Petersen, O. H. *J. Physiol., Lond.* **301**, 295-305 (1980).
8. Schulz, I. & Stolze, H. H. *A. Rev. Physiol.* **42**, 127-156 (1980).
9. Matthews, E. K., Petersen, O. H. & Williams, J. A. *Analyt. Biochem.* **58**, 155-160 (1974).
10. Matthews, E. K., Petersen, O. H. & Williams, J. A. *J. Physiol., Lond.* **234**, 689-701 (1973).
11. Davison, J. S., Pearson, G. T. & Petersen, O. H. in *Biology of Normal and Cancerous Exocrine Pancreatic Cells* (eds Ribet, A., Pradayrol, L. & Susini, C.) 83-88 (Elsevier, Amsterdam, 1980).
12. Petersen, O. H. & Iwatsuki, N. *Ann. N.Y. Acad. Sci.* **307**, 599-617 (1978).
13. Gardner, J. D. A. *Rev. Physiol.* **41**, 55-66 (1979).
14. Burnstock, G. *Pharmac. Rev.* **24**, 509-581 (1972).
15. Burnstock, G. in *Physiological and Regulatory Functions of Adenosine and Adenine Nucleotides* (eds Baer, H. P. & Drummond, G. I.) 3-32 (Raven, New York, 1979).
16. Larsson, L.-I., Fahrenkrug, J., Holst, J. J. & Schaffalitzky de Muckadell, O. B. *Life Sci.* **22**, 773-780 (1978).
17. Sundler, F., Alumets, J., Håkanson, R., Fahrenkrug, J. & Schaffalitzky de Muckadell, O. B. *Histochemistry* **55**, 173-176 (1978).

Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblasts

Chiaho Shih, L. C. Padhy, Mark Murray & Robert A. Weinberg

Department of Biology and Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

We have previously¹ demonstrated that DNA of mouse fibroblasts transformed by 3-methylcholanthrene (3-MC) induced foci of transformed cells when applied to monolayer cultures of NIH3T3 cells, which indicates that at least a part of this phenotype is encoded in DNA sequences. However, our conclusions were confined to the effects of DNAs of 3-MC-transformed mouse fibroblasts on recipient NIH3T3 cells, also of mouse fibroblast origin. To elucidate this phenomenon further, we have prepared DNAs from a series of mouse and non-mouse tumour lines of non-fibroblastic origin and investigated whether tumour transforming genes can act across tissue and species barriers to transform NIH3T3 cells. We find that DNAs obtained from human, rabbit and mouse bladder carcinoma lines, a lung carcinoma line and rat neuroblastoma and mouse glioma lines, are able to induce transformation of NIH3T3 cells on transfection.

DNA prepared from tumours or tumour cell lines were transfected using the calcium phosphate transfection procedure described elsewhere^{1,2}. Each transfected culture was passaged into 12 subcultures and foci of transformants were scored 14–20 days later¹; only foci containing very refractile cells were counted. Because of occasional ambiguities in resolving true transformants from spontaneous overgrowths of the monolayer culture, critical transfections were scored using a double-blind protocol. We observed induction of transformed foci when transfected DNAs were prepared from two classes of tumour cells: neuroblastoma/glioma (Table 1) and carcinoma (Table 2).

Most of the DNAs used in these experiments were prepared from tumour cell lines which had been derived from tumour explants. Each tumour cell line that gave positive results had been shown to possess differentiation markers characteristic of the tumour type from which the cell line had been derived^{3–8}. For example, the neuroblastoma lines had been shown to have excitable membranes, neurite extensions and neurotransmitter enzymes³; tumours induced by cell lines of carcinoma origin possessed keratin and characteristic acinar structures⁵.

As seen in Table 1, foci were induced by the DNAs of one glioma and three neuroblastoma cell lines. The latter were derived from tumours induced in rats *in utero* by exposure to ethylnitrosourea³, and the glioma cell was prepared from a tumour induced by direct implantation of a 3-MC precipitate into the brain of a newborn mouse⁴. The foci produced by these transfections were grown up and their DNA prepared for a second cycle of transfection. The DNAs of these primary foci were all able to induce secondary foci, which demonstrates the serial transmissibility of these transformation-inducing sequences.

Transfection of the DNAs of three different bladder carcinoma cell lines also yielded foci after application to NIH3T3 monolayer cultures as did transfection of DNAs of the mouse Lewis lung carcinoma tumour (Table 2). Again, DNAs prepared from cultures of these primary transfectants were able to induce foci after a second cycle of transfection. The DNAs of a large series of other carcinoma cell lines yielded few or no foci on transfection (Table 2).

When the cells of foci resulting from transfections induced by neuroblastoma, glioma and carcinoma DNAs were grown into mass cultures and seeded subcutaneously into newborn NFS mice, rapidly growing fibrosarcoma-like tumours were observed

after 2 weeks with almost 100% incidence (Tables 1 and 2). Thus, the transfected DNA induced foci of cells which were highly tumorigenic. In contrast, normal, untransfected NIH3T3 cells resulted in a low incidence of tumours (1/7, 1/10) which only appeared 4 weeks after subcutaneous injection into the newborn mice (Tables 1, 2).

All attempts at rescue of C-type transforming retroviruses from the transfected cells were unsuccessful. We also attempted rescue of virus from cells transformed by Moloney sarcoma virus and observed the release of high titres of focus-forming virus. We conclude that transmissible C-type viruses are not responsible for the phenomenon of focus induction in donor or recipient cells.

Human donor DNA was used in one of these transfections to obtain direct biochemical proof that the transfected cells had taken up and stabilized human DNA. This depended on the fact that the few cells which acquire a selected marker gene also acquire a large number of other donor DNA fragments which enter the cells during the same transfection event⁹. In contrast, other cells in the culture which have not acquired the selected marker gene do not show this array of co-transfected donor sequences⁹.

In the present study, cells which have acquired the human bladder carcinoma transformation gene should also contain a large array of other co-transfected human DNA fragments, many of which should bear a copy of a highly repeated (300,000 per haploid) sequence which is interspersed throughout the human genome¹⁰. This sequence is not detectable in normal mouse DNA. Such sequences can be detected by the Southern blot procedure¹¹ using a cloned probe of the highly repeated human sequence^{10,12}. Figure 1, lanes c–g, reveals the presence of many segments bearing the human repeated sequence in the DNAs of five different cell lines, each derived from a separate focus induced by human bladder carcinoma DNA. DNAs not derived from *bona fide* transfectants show no trace of these acquired human sequences (lanes a, b). The successfully transformed cells are thus descended from the small subpopulation of cells in the initially transfected culture⁹ which was able to take up and fix large amounts of donor DNA. We conclude that the transformation of these cells was a direct consequence of the acquisition of sequences of a human bladder carcinoma DNA. This or similar hybridization tests may be used as assays to distinguish *bona fide* transfectants from other

Table 1 Transfection of DNAs of glioma and neuroblastoma cell lines

Donor DNA	Tissue type, strain and species	Mode of tumour induction	Efficiency of primary transfection (FFU per 75 µg DNA)	Efficiency of secondary transfection (FFU per 75 µg DNA)	Tumorigenicity of primary transfectant	Retrovirus rescue assay of primary transfectant
G26-20	C57BL/6 mouse glial*	3-methylcholanthrene precipitate was implanted directly into mouse brain	10†	16†, 22†	ND	ND
NIH3T3	NIH/Swiss embryo		0†		1/10‡, 1/7‡	0
B50	BDIX rat, neurone§	Fetal rats were exposed to nitrosoethylurea transplacentally	6, 9, 13, 16	78, 82†, 100†	8/8	0
B104	Same		9, 31	500	6/6, 7/7	0
B103	Same	by injection into pregnant	93, 122†	12†, 12†	9/9	0
B65	Same	mother on day 15 after	0, 0–1, 0†			
B15	BDIX rat, glial	conception	0†, 0†, 1			

One transfection (fourth and fifth columns) is defined as 75 µg DNA applied to 1.5×10^6 NIH3T3 cells and values represent total number of foci per transfection. Transfection procedures were as described elsewhere¹. Tumorigenicity of transfectant clones was determined by injecting subcutaneously $1-2 \times 10^6$ cells suspended in 0.1 ml PBS into 1–2-day-old newborn mice of NFS or NIH/Swiss outbred strains. Tumour formation was seen after 2 weeks. NIH3T3 cells were injected in an identical manner as a control. Retrovirus rescue assays were designed to induce the release of any defective, transforming viruses which might be present in the transfected cultures as described elsewhere¹. The titre of helper virus released from infected cultures in these rescue assays was routinely 1×10^7 plaque forming units per ml. Infection of Moloney murine sarcoma virus-transformed NIH3T3 non-producer cells with helper virus, performed in parallel as positive control, resulted in release of $>5 \times 10^3$ focus forming units (FFU) of transforming virus per ml. Lines B65 and B15 were used as negative controls for the transfection assay. ND, not determined. IMR32 human neuroblastoma¹³ and NB2A (c1300) mouse neuroblastoma¹⁴ DNAs did not induce a significant number of foci.

* See ref. 15 for origin.

† Not performed in double-blind.

‡ These tumours were seen only at 4 weeks post-inoculation.

§ See ref. 3 for origin.

Table 2 Transfection of DNAs of carcinoma and other tumour lines

Donor DNA	Tissue type, strain and species	Mode of tumour induction	Efficiency of primary transfection (FFU per 75 µg DNA)	Efficiency of secondary transfection (FFU per 75 µg DNA)	Tumorigenicity of primary transfectants	Retrovirus rescue assay of primary transfectant
Lewis lung carcinoma tumour	C57BL/6 mouse lung epithelial*	Spontaneously occurring in animal	7†, 12, 13, 66	95, 105†	9/9	0
Liver of Lewis lung carcinoma-bearing mice	C57BL/6 mouse liver		1-2, 4			
MB48BT-26	C57BL/6 mouse bladder epithelial‡	Isolated bladder epithelial explant was exposed to 7, 12-dimethylbenzanthracene <i>in vitro</i>	48, 56†	10†	8/8	0
NIH3T3	NIH/Swiss mouse embryo fibroblast		0, 0, 0-1		0/9, 0/6	
RBC	NZW rabbit bladder, epithelial§	As above, except benzo-a-pyrene was used	26, 101, 165	52	7/7	0
Rabbit bladder	NZW rabbit bladder		1, 1			
EJ	Human bladder, epithelial	Spontaneously occurring	40†, 66†, 70†	83†, 100†	9/9, 15/15	0
Human epidermal cell strain	Human epidermis		0†, 0†			

DNA transfection procedures were as described elsewhere¹; one transfection is defined as 75 µg DNA for 1.5×10^6 NIH3T3 recipients. Tumorigenicity and retrovirus rescue assay were carried out as described in Table 1 legend. The DNAs of the following tumour cell lines induced few or no foci in a transfection assay: SCC9, SCC12X, SCC4 squamous cell carcinoma cell lines¹⁶; A1663 a human bladder carcinoma line¹⁷; A549, a human lung carcinoma line¹⁷; Detroit 562, a human pharyngeal carcinoma line¹⁸; RT4, a human bladder carcinoma line¹⁹; BC6 a rat bladder carcinoma line (obtained from Dr Howard Green, MIT); and Nettesheim squamous cell carcinoma (obtained from A. D. Little Co.). The DNAs of other tumour cell lines induced few or no foci in a DNA transfection assay: HTC rat hepatoma established by Dr Gordon Tomkins; 7777 rat hepatoma²⁰; human Murawski melanoma and human Wolfe melanoma (obtained from Dr Lloyd Old); A375 human melanoma¹⁷; M3 mouse melanoma²¹ and B16 mouse melanoma²².

*For information on origin of cell line see Dr M. R. Lewis, unpublished data.

† Not performed in double-blind.

‡ For origin of cell line see ref. 6.

§ See ref. 5.

|| Origin of cell line Dr J. Daly, unpublished data.

colonies in cultures which have become transformed by mechanisms other than transfection.

The positive results with the human, rabbit and rat tumour DNAs indicate that tumour transforming genes are able to act across species barriers—they may also act across tissue barriers: for example, transforming sequences of neuroblasts and epithelial cells can transform the recipient fibroblasts. In addition, transforming sequences that are active in these transfections (Tables 1, 2) may be derived from tumours induced by a variety of carcinogens which have been administered in different ways. We conclude that the transfection assay can reveal transformation sequences in tumours of diverse origin and is therefore not limited to analysis of donor tumour DNAs of mouse fibroblast origin. The demonstration of these transmissible genes from various tumours will allow their molecular cloning

and the analysis of the molecular basis of non-viral carcinogenesis.

We thank the following people for providing tumour cell lines: Drs I. Summerhayes, L. B. Chen, J. Rheinwald, H. Green, O. Kehinde, J. Patrick, T. Yean-Kai, I. Wodinsky, G. Todaro,

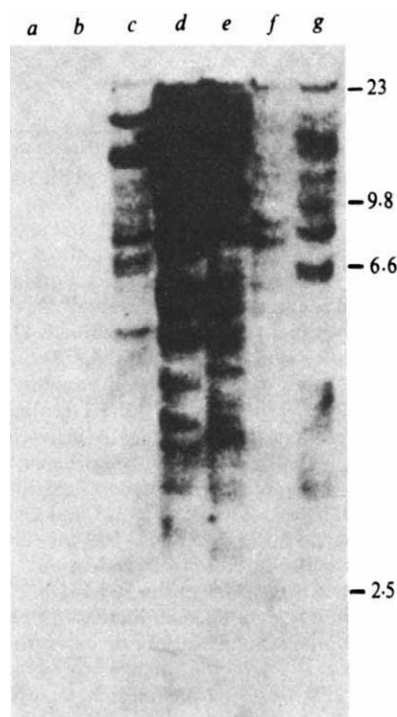


Fig. 1 Detection of human DNA sequences in transformed NIH3T3 cells following transfection with EJ carcinoma DNA. DNA was prepared from five foci; 10 µg of each DNA was digested with *EcoRI* and electrophoresed through a 1% agarose gel in 40 mM Tris (pH 7.9), 50 mM sodium acetate and 1 mM EDTA. The gel was run at 1.2 V cm^{-1} for 16 h. After electrophoresis, the DNA was transferred to nitrocellulose by the method of Southern¹¹, and the resulting blot hybridized with the BLUR8 clone of the human repetitive 'Alu family' DNA¹². The Alu probe was ³²P-labelled by nick translation²³ and 5×10^6 c.p.m. of labelled probe was hybridized to the blot in the presence of 10% dextran sulphate²⁴. Hybridization was carried out for 40 h followed by a 12-h exposure to X-ray film. Whole cellular human DNA used as a positive control gave extensive blackening throughout the entire lane, without discernible bands (not shown here). Approximate size ranges (in kilobases) are indicated to the right of the gel. a, b, DNAs of untransformed NIH3T3 cells, picked and cloned from dishes which also contained transformed foci. c-g, DNAs from different transformed foci of NIH3T3 cells, derived from transfection of EJ donor DNA.

L. Old, J. Fogh, T-W. Chang and C. Cepko; Drs P. Deininger, C. Schmid and W. Jelinek for the BLUR8 clone; Dr James Murphy for histology of several tumour samples. We also thank A. Dannenberg and H-W. Hsu for technical assistance and Dr B. Shilo for useful discussions. This work was supported by funds from the Rita Allen Foundation of New York and the NIH, including core grant CA14051 and program project CA26717. C.S. was supported by a training grant to the MIT Department of Biology, L.C.P. was the recipient of an Eleanor Roosevelt International Fellow and M.M. was a fellow of the Damon Runyon-Walter Winchell Foundation.

Received 22 September 1980; accepted 13 January 1981.

- Shih, C., Shilo, B., Goldfarb, M. P., Dannenberg, A. & Weinberg, R. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5714 (1979).
- Graham, F. L. & van der Eb, A. J. *Virology* **52**, 456 (1973).
- Schubert, D. *et al. Nature* **249**, 224 (1974).
- Dawson, C. T., Sundarraj, N. & Pfeiffer, S. E. *J. biol. Chem.* **252**, 2777 (1977).
- Summerhayes, I. C., Sun, T. T. & Chen, L. B. *J. Cell Biol.* (in the press).
- Summerhayes, I. C. & Franks, L. M. *J. natn. Cancer Inst.* **62**, 1017 (1979).
- Marshall, C. J., Franks, L. M. & Carbonell, A. W. *J. natn. Cancer Inst.* **58**, 1743 (1977).
- Sugiura, K. & Stock, C. C. *Cancer Res.* **15**, 38 (1955).
- Wigler, M. *et al. Cell* **16**, 777 (1979).
- Rubin, C. M., Houck, C. M., Deininger, P. L., Friedmann, T. & Schmid, C. W. *Nature* **284**, 372 (1980).
- Southern, E. M. *J. molec. Biol.* **98**, 503 (1975).
- Jelinek, W. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 1398 (1980).
- Tumilowicz, J. J., Nichols, W. W., Cholon, J. J. & Green, A. E. *Cancer Res.* **30**, 2110 (1970).
- Klebe, R. J. & Ruddle, F. H. *J. Cell Biol.* **43**, 69 (1969).
- Sugiura, K. in *Methods of Animal Experimentation* Vol. 2 (ed. Gay, W. L.) 171-222, (Academic, New York, 1969).
- Rheinwald, J. G. & Beckett, M. A. *Cell* **22**, 629 (1980).
- Giard, D. J. *et al. J. natn. Cancer Inst.* **51**, 1417 (1973).
- Peterson, W. D. Jr, Stulberg, C. S. & Simpson, W. F. *Proc. Soc. exp. Biol. Med.* **136**, 1187 (1971).
- Rigby, C. C. & Franks, L. M. *Br. J. Cancer* **24**, 746 (1970).
- McMahon, J. B., Kelleher, P. C. & Smith, C. J. *Biochem. biophys. Res. Commun.* **76**, 1144 (1977).
- Yasumura, Y., Tashjian, A. H. & Sato, G. H. *Science* **154**, 1186 (1966).
- Handbook on Genetically Standardized Jax Mice* (Jackson Memorial Laboratory, Bar Harbor, 1962).
- Rigby, P. W. J., Dickmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237 (1977).
- Wahl, G., Stern, M. & Stark, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3683 (1979).

Identification of mutations affecting replication control of plasmid Clo DF13

Antoine R. Stuitje, Cornelis E. Spelt, Eduard Veltkamp & H. John J. Nijkamp

Department of Genetics, Biological Laboratory, Vrije Universiteit, De Boelelaan 1087, 1007 MC Amsterdam, The Netherlands

The bacteriocinogenic plasmid Clo DF13, originally isolated from *Enterobacter cloacae*¹, is stably maintained in *Escherichia coli*² to the extent of about 10 copies per cell³. Its replication⁴ resembles that of many other small, multicopy plasmids⁵⁻⁸; plasmid-encoded protein is not required⁹⁻¹² but plasmid-specific genetic information is involved in regulation of replication as both conditional¹³ and nonconditional³ copy-number mutants of Clo DF13, and transcomplementable^{14,15} copy-number mutants of plasmid Col E1 have been described. The sequences essential for replication of Col E1 (refs 16, 17) and Clo DF13 (refs 18, 19) have been identified within a region surrounding the replication origin. Initiation of Col E1 replication is preceded by transcription of the origin region, providing the RNA primer at the origin²⁰. However, transcription in the opposite direction results in a small transcript of ~100 nucleotides (RNA-100) for both Col E1 (refs 21, 22) and Clo DF13 (ref. 23). Data suggest that Col E1 RNA-100 acts as a negative control element for the initiation of replication^{20,24}. We show here that single base transitions in the RNA-100 cistron of Clo DF13 can result in a nonconditional increase in plasmid copy-number. Also, sequence analysis has revealed that a specific base transition in a DNA region, apparently involved in both termination and initiation of transcription towards the replication origin, results in a thermosensitive plasmid copy-number.

Both nonconditional³ and conditional¹³ replication-control mutants of plasmid Clo DF13 (8.6 kilobases) have been isolated after chemical mutagenesis, and characterized^{13,25}. As reported elsewhere¹⁸, the Clo DF13-*cop3* mutation (Table 1) was mapped by insertion of transposons and successive deletion of specific restriction fragments. Similar studies on the Clo DF13-*cop1* (Ts) and the Clo DF13-*cop2* mutants (Table 1) show that these mutations are also located within the region (1.8-15%) surrounding the replication origin. These copy mutations were localized more precisely within this region by *in vitro* recombination experiments (see Fig. 1 legend). Subsequently, specific regions of the different copy mutants were sequenced as indicated in Table 1. By comparing these sequences, the complete nucleotide sequence between the *Hae*II site at 1.8% and the *Bam*HI site at 15% was deduced (Fig. 2). Furthermore these studies show that these different *cop* mutations are, in fact, point mutations, and all represent base transitions (Fig. 2).

Recently, Itoh and Tomizawa²⁰ demonstrated that the initiation of DNA replication of plasmid Col E1 *in vitro* is preceded by transcription of the origin region, towards the origin of replication, generating a precursor primer RNA molecule (Fig. 1). Furthermore, they showed that when transcription passes the origin, the RNA transcript is able to form a stable hybrid with this region. In the next step RNase H specifically cleaves the hybridized RNA, producing the RNA-DNA hybrid with a free 3' OH end for the initiation of DNA replication by DNA polymerase I. The 5' end of the 'pre-primer' RNA has been sequenced and the corresponding region within the sequence of plasmid Clo DF13 is indicated in Fig. 2.

Figures 2 and 3 show that the Clo DF13 *cop1* (Ts) mutation involves a base transition at position 720. The region surrounding this mutation has several interesting features: first, the nucleotide sequence of this region (see Fig. 4b) resembles that of regions known to be involved in termination of transcription²⁶. A similar role for the region in Clo DF13 is supported by data²³ which shows that transcription from the cloacin promoter results in the synthesis of two RNA molecules of 2.2 and 2.4 kilobases respectively. The terminator for the transcription resulting in the synthesis of the 2.4-kilobase RNA was located at 9.3%, and therefore maps within the sequence mentioned above (T2, Fig. 1). Second, the nucleotide sequence of the region surrounding the *cop1* (Ts) mutation (position 696-735) is completely conserved when compared with the corresponding sequence of plasmid Col E1 (ref. 16). The position of the initiation site of transcription of the Col E1 pre-primer RNA, as determined by Itoh and Tomizawa²⁰, implicates the Col E1 region mentioned above in promotion of the synthesis of this RNA. Therefore, the corresponding region of plasmid Clo DF13 which surrounds the *cop1* (Ts) mutation might also act as promoter for RNA synthesis (P3, Fig. 1). In view of these data, the *cop1* (Ts) mutation could be a promoter-type or terminator-type mutation. Alternatively, it could be a mutation at the site of action of a control element. Further experiments are in progress to discriminate between these possibilities and determine whether the effect of the *cop1* (Ts) mutation, a temperature-inducible increase of the plasmid copy-number, is caused by an increased synthesis of pre-primer RNA at the elevated temperature.

The nonconditional *cop2* and *cop3* mutations are located within a Clo DF13 region which is transcribed *in vivo* and *in vitro* resulting in a RNA molecule of ~100 nucleotides^{19,23} (RNA-100, see Figs 2, 3). The corresponding region of plasmid Col E1 is also transcribed *in vitro*^{21,22}, producing a similar molecule. The nucleotide sequence of the latter RNA-100 species has been determined and is indicated in Fig. 4a. Perusal of the Clo DF13 and Col E1 sequences encoding this RNA-100 species strongly suggest that these RNA molecules are not translated. Furthermore, it was observed that although the primary structures of the Clo DF13 and Col E1 RNA-100 species differ largely, the secondary structure seems to be highly conserved (Fig. 4a). The position of the *cop2* and *cop3* mutations within the sequence (Fig. 2) makes it unlikely that these mutations directly affect the synthesis of either the RNA-100 or the pre-primer RNA.

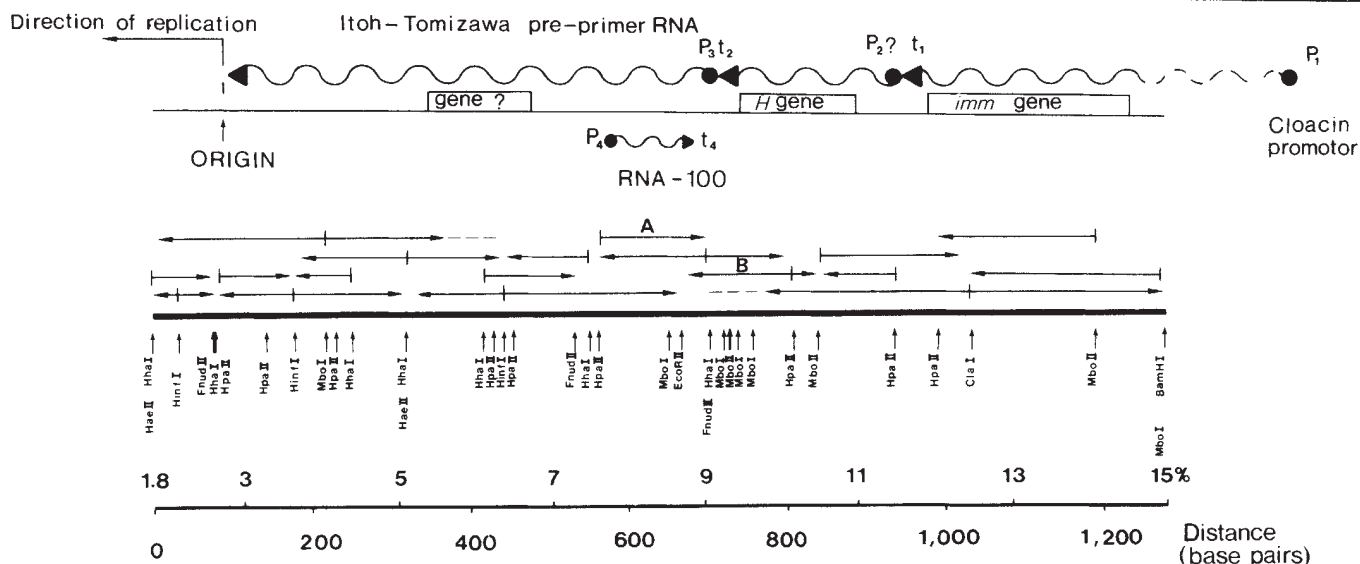


Fig. 1 Detailed physical/genetic map of the region between the *HaeII* site and the *BamHI* site at respectively 1.8 and 15% on the Clo DF13 map¹⁹. The positions of the various restriction endonuclease cleavage sites were determined as reported previously¹⁹ and verified by sequence analysis. The exact position of the immunity gene (*imm*) could be determined by DNA sequence analysis of this region, as part of the amino acid sequence of the Clo DF13 immunity protein has been elucidated²⁷. The region of the Clo DF13 plasmid genome encoding the H protein was previously described²³. The nucleotide sequence probably encoding this polypeptide of about 50 amino acids is indicated in Fig. 2. The region possibly encoding an arginine-rich, basic polypeptide of about 45 amino acids, for both Col E1 and Clo DF13 (ref. 19), is indicated as 'gene?'. As reported²³, transcription initiated at the cloacin promoter (*P*₁) terminates at *t*₁ or *t*₂ resulting in RNA molecules of ~2,200 and 2,400 nucleotides respectively. By comparing the Clo DF13 and Col E1 origin sequences, the position of the Clo DF13 RNA-100 cistron was deduced¹⁹. As indicated, the origin of replication maps within the 1.8–5% *HaeII* restriction fragment²⁸. By partial cleavage of a Clo DF13-*cop3* derivative (pEV26)¹⁸ with *HaeII* endonuclease in the presence of *S*₁ nuclease, followed by ligation with *T*₄ ligase and transformation of *E. coli*, we have isolated a plasmid which has an altered *HaeII* (1.8–5%) 'origin' fragment and still contains the *cop3* mutation (A.R.S. *et al.*, unpublished results). Recombinant plasmids containing both wild-type (wt) and *cop3* *HaeII* fragments were obtained by transforming cells to ampicillin resistance with a ligated mixture of wild-type and *cop3* *HaeII* fragments. The composition of the recombinant plasmids was easily determined from the *HaeII* digestion patterns. The Cop phenotype was deduced from the minimal inhibitory concentration of ampicillin of cells harbouring these plasmids. By these types of study the Clo DF13-*cop3* and -*cop2* mutations were localized more precisely within the 1.8–15% origin region (see Table 1).

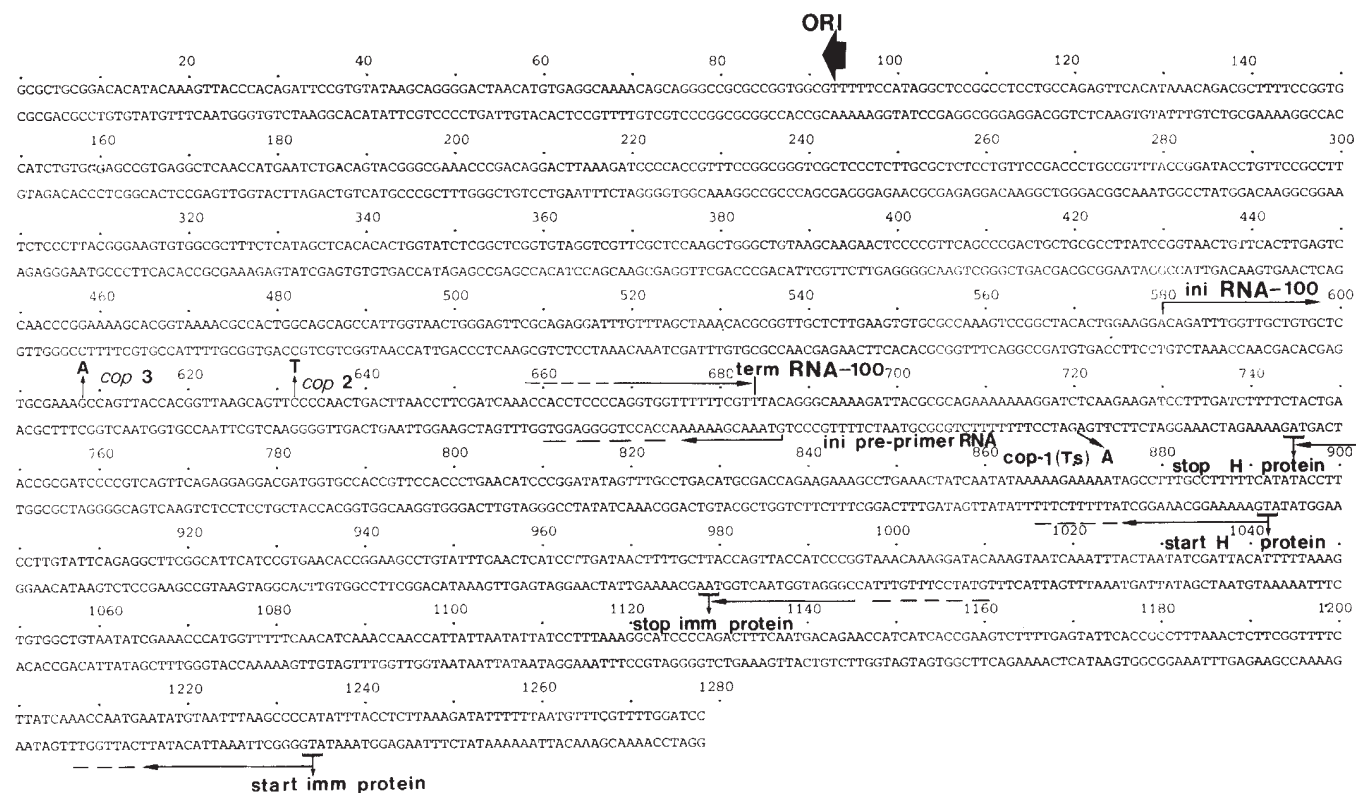


Fig. 2 The nucleotide sequence of the Clo DF13 region between the *HaeII* site at 1.8% and the *BamHI* site at 15% on the Clo DF13 physical map (see also Fig. 1). The sequencing procedures were as described previously¹⁹. The sequence strategy is indicated in Fig. 1 legend; the arrows in Fig. 1 show the part of the restriction fragments sequenced. The direction of the arrows indicates the strands which were sequenced, aligned in 5' → 3' direction (tail to head). Start and stop indicate respectively the start and stop codons for proteins; 'ini' and 'term' represent the 5' and 3' ends of RNAs respectively.

Table 1 Properties of Clo DF13 and mutant plasmids

Plasmid	Derivation	Relevant properties	Copies per cell	MW (kilobases)	Clo DF13 region present (%)	Map* position of the <i>cop</i> mutation (%)	Clo DF13 region of plasmid sequence (%)	Ref. source
Clo DF13	Naturally occurring in <i>E. cloacae</i>	Cloacin DF13	10	8.6	0–100	—	—	1
Clo DF13- <i>cop3</i>	Clo DF13	Cloacin DF13, nonconditional copy mutant	70	8.6	0–100	5–11.5	—	21
Clo DF13- <i>cop2</i>	pEV5	Nonconditional copy mutant, Ap	200	2.9	1.8–15	5–15	5–15	Stuitje (unpublished)
Clo DF13- <i>cop1</i> (Ts)	pJN03	Cloacin DF13, ApTs-copy mutant Ts-kill	Temperature dependence	13.1	0–100	1.8–9.5		14
pEV5	Clo DF13::Tn901 insertion mutant	Ap	10	2.9	1.8–15		5–15	Stuitje (unpublished)
pEV22	Clo DF13- <i>cop3</i> ::Tn901 insertion mutant	Ap	200	2.8	1.8–11.5		1.8–11.5	4
pVU206	Clo DF13- <i>cop1</i> (Ts)	Ap Ts-copy mutant Ts-kill	Temperature dependence	3.3	1.8–15 27–29		1.8–15	Stuitje (unpublished)

Ts, temperature sensitive.

* The different *cop* mutations were localized more precisely within the 1.8–15% origin region by recombining the *Hae*II 'ori' fragments of the different mutant plasmids, as described in Fig. 1 legend.

Although our data on the *cop2* and *cop3* mutations do not exclude a possible role of the RNA-100 as a positive control element, they are consistent with previous findings that favour the RNA-100 as a negative control element for the initiation of DNA replication^{20,24}. Furthermore, results²⁵ suggest that the RNA-100 is involved in plasmid incompatibility, a plasmid function that seems to be intimately connected with replication control. In fact, it has been reported²⁰ that the presence of the RNA-100 species of plasmids belonging to the Col E1 incompatibility group inhibits the formation of the primer *in vitro*. The sequence data described in this letter indicate that there are at least two distinct regions that are important in the control of Clo DF13 replication. The first encodes a RNA-100 molecule whose synthesis is initiated ~490 nucleotides upstream from the replication origin, whereas the second, which is located about 620 nucleotides upstream from the origin,

seems to be involved in both termination and initiation of transcription towards the origin. The identification of the different mutations inevitably evokes questions on the mechanism of replication control of small plasmids such as Col E1, for example: (1) is the frequency of initiation of DNA replication controlled at the level of pre-primer RNA synthesis; (2) is the synthesis of pre-primer RNA regulated at P3/t2 (see Fig. 1); (3) in which way does the RNA-100 species control plasmid replication and are, in addition to this RNA, other plasmid-encoded gene products involved? Studies on additional copy-control mutants, including complementation experiments between the mutants and studies on the effects of alterations generated *in vitro* by substitution or deletion of specific sequences at different regulatory sites, may elucidate the actual mechanism of plasmid DNA replication control.

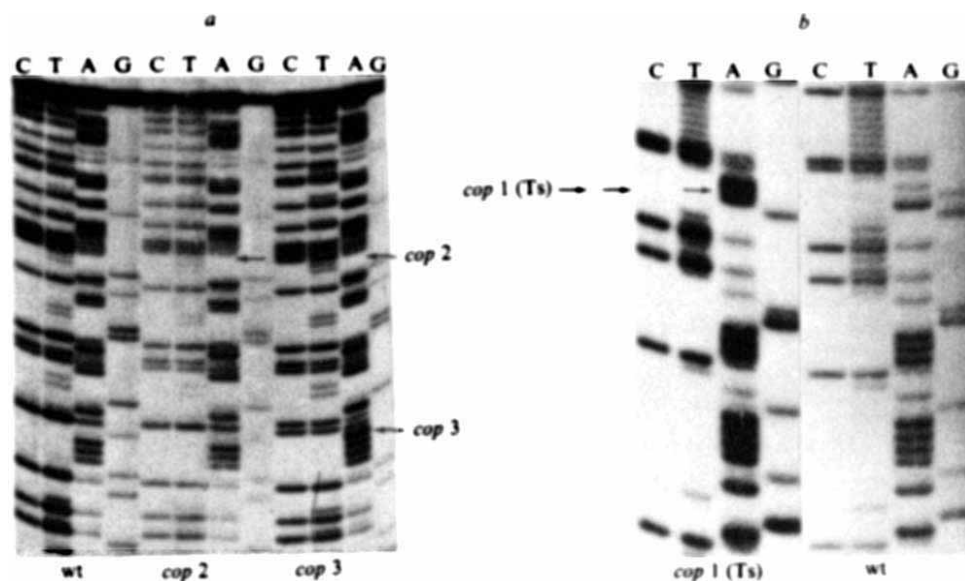


Fig. 3 *a*, Result of the Gilbert-Maxam sequencing procedure applied to fragment A (labelled at the 5' end of the *Hpa*II site as described previously¹⁹; see Fig. 1) derived from the wild-type (wt) Clo DF13 and the *cop2* and *cop3* mutants. Note the depression in the wild-type sequence near the position of the *cop3* mutation. The actual wild-type sequence 5'-A-A-A-G-C-C-A-3', instead of 5'-A-A-A-G-C-A-3', was deduced by sequence analysis of the complementary strand. *b*, Result of the Maxam-Gilbert sequencing procedure applied to fragment B (5' end-labelled at the *Hpa*II site; see Fig. 1) of the wild-type Clo DF13 plasmid and the Clo DF13-*cop1* (Ts) mutant.

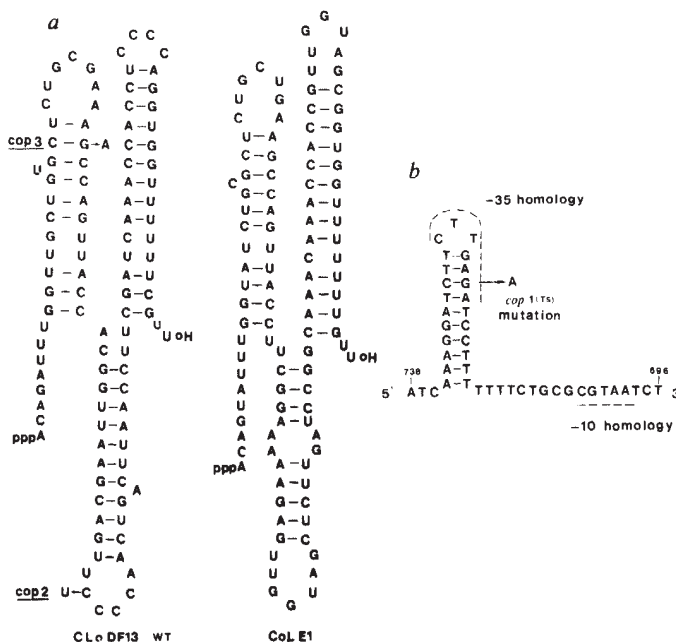


Fig. 4 *a*, Comparison of the secondary structures of the RNA-100 species encoded by the plasmids Col E1 and Clo DF13. The resident *cop2* (C → U) and *cop3* (G → A) mutations are indicated. The most stable secondary structure was estimated by the procedure described by Tinoco *et al.*²⁹ and the ΔG (25 °C) values estimated for the wild-type (wt) Clo DF13, *cop2*, *cop3* and Col E1: 100-nucleotide RNAs are respectively 50.4, -50.2, -41.2 and -45.6 kcal. *b*, Secondary structure of the DNA region surrounding the *cop 1* (Ts) mutation (G → A transition). This region is presumed to be involved in both termination and initiation of transcription towards the origin (transcription proceeds from left to right). The sequences which are supposed to be involved in recognition and binding of RNA polymerase are indicated by -35 and -10 homology respectively²⁶.

This study was supported in part by The Netherlands Foundation for Chemical Research (SON) with financial aid from The Netherlands Organization for Advancement of Pure Science.

Received 10 October 1980; accepted 2 January 1981.

1. Tietze, G. A., Stouthamer, A. H., Jansz, H. S., Zandberg, J. & van Bruggen, E. F. *J. Molec. gen. Genet.* **106**, 48-65 (1969).
2. Kool, A. J., Pranger, M. & Nijkamp, H. J. *J. Molec. gen. Genet.* **115**, 314-323 (1972).
3. Kool, A. J. & Nijkamp, H. J. *J. Bact.* **120**, 569-578 (1974).
4. Veltkamp, E. & Stuitje, A. *R. Plasmid* (in the press).
5. Staudenbauer, W. L. *Curr. Topics Microbiol. Immun.* **83**, 93-157 (1978).
6. Bolivar, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **74**, 5265-5269 (1977).
7. Conrad, S. E., Wold, M. & Campbell, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 736-740 (1979).
8. Chang, A. C. Y. & Cohen, S. N. *J. Bact.* **134**, 1141-1156 (1978).
9. Kahn, M. & Helinsky, D. R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2200-2204 (1978).
10. Donoghue, D. J. & Sharp, P. A. *J. Bact.* **133**, 1287-1294 (1978).
11. Tomizawa, J. *Nature* **257**, 253-254 (1975).
12. Staudenbauer, W. L. *Molec. gen. Genet.* **145**, 273-280 (1976).
13. Andreoli, P. M., Brandsma, R. A., Veltkamp, E. & Nijkamp, H. J. *J. Bact.* **135**, 612-621 (1978).
14. Twigg, A. J. & Sherratt, D. *Nature* **283**, 216-218 (1980).
15. Shepard, H. M., Gelfand, D. H. & Polisky, H. *Cell* **18**, 267-275 (1979).
16. Oka, A. *et al. Molec. gen. Genet.* **172**, 151-159 (1979).
17. Backman, K., Betlach, M., Boyer, H. W. & Yanowski, S. *Cold Spring Harb. Symp. quant. Biol.* **43**, 69-76 (1979).
18. Stuitje, A. R., Veltkamp, E., van den Elzen, P. J. M. & Nijkamp, H. J. *J. Nucleic Acids Res.* **5**, 1801-1820 (1978).
19. Stuitje, A. R., Veltkamp, E., Maat, J. & Heijneker, H. L. *Nucleic Acids Res.* **8**, 1459-1473 (1980).
20. Itoh, T. & Tomizawa, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2450-2454 (1980).
21. Patient, R. K. *Nucleic Acids Res.* **6**, 2647-2666 (1979).
22. Morita, M. & Oka, A. *Eur. J. Biochem.* **97**, 435-443 (1979).
23. Van den Elzen, P. J. M., Konings, R. N. H., Veltkamp, E. & Nijkamp, H. J. *J. Bact.* **144**, 579-591 (1980).
24. Conrad, S. E. & Campbell, J. L. *Cell* **18**, 61-71 (1979).
25. Hashimoto-Gotoh, T. & Inselberg, J. *J. Bact.* **139**, 608-619 (1979).
26. Rosenberg, M. & Court, D. A. *Rev. Genet.* **13**, 319-353 (1979).
27. Van den Elzen, P. J. M. *et al. Nucleic Acids Res.* (in the press).
28. Stuitje, A. R., Veltkamp, E., Weyers, P. J. & Nijkamp, H. J. *J. Nucleic Acids Res.* **6**, 71-80 (1979).
29. Tinoco, I. *et al. Nature new Biol.* **246**, 44, 40-41 (1973).

Differential usage of iso-accepting tRNA^{Ser} species in silk glands of *Bombyx mori*

Danièle Hentzen, Anna Chevallier & Jean-Pierre Garel

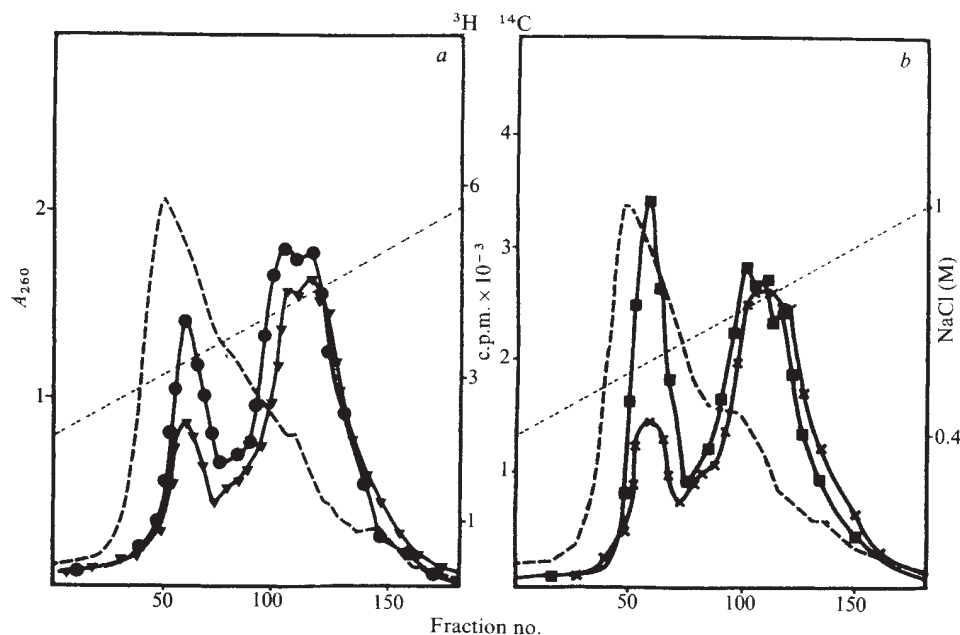
Department of General and Applied Biology, Université Claude Bernard, Lyon I, 43, Bd du 11 Novembre 1918, F-69622 Villeurbanne Cedex, France

The rapid development of the silk glands of *Bombyx mori* during the last larval instar shows two phases. During the first 4 days, in both the middle and posterior parts of the silk glands, the ribosomal machinery is assembled and the synthesis of house-keeping proteins starts. During the second phase (the last 4 days), the middle part of the gland synthesizes ~45 mg of the silk protein sericin (31% serine) and the posterior part of the gland synthesizes ~130 mg of the silk protein fibroin (46% glycine, 29% alanine and 12% serine). Silk fibroin and sericin are detectable by the second day and represent 80 and 50% respectively of the total proteins produced at day 8 (refs 1-4). It is known that the tRNA population of the posterior part of the gland is quantitatively adapted to fibroin codon frequency during this period⁵⁻⁹ but little is known about the situation in the middle part except for the observation that it contains more tRNA^{Ser} than does the posterior part. We show here that the two parts contain, and presumably use, different iso-accepting species of tRNA^{Ser}, the middle part using tRNA₁^{Ser}, which recognizes AGU and AGC codons, and the posterior part using tRNA₂^{Ser} which recognizes UCA. We also suggest that this differential adaptation of the tRNA^{Ser} species is under transcriptional control as the two species are accumulated at different rates, but degraded at the same rate.

tRNA^{Ser} species were fractionated and quantitated using benzoylated DEAE (BD)-cellulose chromatography¹⁰. Figure 1 shows chromatographic profiles of acylated Ser-tRNA from middle and posterior silk glands of *B. mori* at two different stages of development: the growth phase (day 2) and at the end of silk secretion (day 8). A striking difference occurs in the distribution of the subsets of tRNA^{Ser} species: in the fibroin-producing part, the tRNA₂^{Ser} group is predominant (82% during the secretion phase) but in the sericin-producing compartment, the tRNA₁^{Ser}/tRNA₂^{Ser} ratio changes from 3.1 at the start of the growth phase to 1.8 during the secretion phase. The tRNA₁^{Ser} species then represents 36% of the iso-tRNA^{Ser} family. From base composition, codon recognition and partial sequencing, we know that the tRNA₁^{Ser} species is capable of translating AGU and AGC codons, whereas the tRNA₂^{Ser} group recognizes UCA codons^{6,11,12}.

Refinements were made using two-dimensional polyacrylamide gel electrophoresis^{13,14}. As shown in Fig. 2, five spots can be assigned to iso-tRNA^{Ser} species. The tRNA₂^{Ser} subset fractionates according to the diagonal into three species: two tRNA_{2a}^{Ser} subspecies with the lowest mobility and one tRNA_{2b}^{Ser} with the highest mobility (Fig. 2a); tRNA₁^{Ser} migrates with an intermediate mobility. It seems to be heterogeneous and resolves into two distinct spots (Fig. 2b-d). Cross-identifications using purified tRNA^{Ser} species collected from a specifically retarded elution on BD-cellulose chromatography, done in the absence of Mg²⁺ ions¹⁵, yield the final pattern shown in Fig. 2c, d. The differential accumulation of iso-tRNA^{Ser} species is made clear on the electrophoretic maps in Fig. 3. It can be seen that the tRNA₁^{Ser} species is more abundant in the middle silk gland (MSG) than in the posterior silk gland (PSG) during the silk secretion phase. Polyacrylamide electrophoresis corroborates the BD-cellulose chromatographic profiles. An unexpected result was the unusual composition of the tRNA₂^{Ser} subset. tRNA_{2b}^{Ser} carrying the IGA anticodon is usually the major

Fig. 1 Chromatography on BD-cellulose of Ser-tRNA from posterior (PSG) and middle (MSG) silk glands of *B. mori* silkworm. *a*, Elution profiles of tRNA extracted from the middle part on the second day (growth phase) of the fifth larval instar and acylated with ^3H -serine ($\bullet$), and tRNA from the posterior part acylated with ^{14}C -serine (∇). The co-chromatography was achieved on a 40×2.2 cm BD-cellulose column eluted with a linear gradient of 0.2–1.0 M NaCl in 2 l of buffer: 10 mM sodium acetate, pH 4.7, 10 mM MgCl_2 at 25°C . 10-ml fractions were treated with 0.1 volume of 55% trichloroacetic acid in the presence of 0.5 mg carrier tRNA. The precipitate was filtered on to a glass fibre disk and radioactivity measured in a liquid scintillation photometer. This experiment was repeated twice after reversing the labelling. Absorbance was measured at 260 nm (---). *b*, Similar experiments with tRNA extracted on the eighth day of the fifth instar: ($\times$) PSG tRNA acylated with ^3H -serine; ($\blacksquare$) MSG tRNA acylated with ^{14}C -serine.



species among the eukaryotic tRNA^{Ser} family able to decode UCU, UCC and, with less efficiency, UCA; a minor tRNA^{Ser} species is able to recognize the rare UCG codon^{15,16}. The tRNA^{Ser} species, which accumulates at a higher level in both silk gland parts than in other non-silk producing tissues (see the carcass in Fig. 3c), might be able to decode specifically the UCA codon—the major serine codon found in fibroin mRNA¹⁷.

A side conclusion can be drawn from these results, based on differential accumulation of specific iso-tRNA^{Ser} species related to the developmental stage of a tissue. The quantitative adaptation of tRNA species to codon frequency demonstrated for a specific and typical mRNA—fibroin or globin¹⁶—leads us to

suggest that the codon usage for serine in the sericin message is somewhat different from that found in fibroin mRNA. The high level of tRNA^{Ser} in the MSG can be easily explained if we state that sericin mRNA carries AGU and/or AGC codons in addition to the UCA codon.

The specific requirement of the iso-tRNA^{Ser} family in decoding different mRNAs in two distinct silk gland parts remains to be understood. We have studied tRNA synthesis using *in vivo* short pulse-labellings of tRNA throughout the last larval instar and then analysed the tRNA species on electrophoretic maps. Synthesis rates of tRNA are not uniform: rapid changes occur, which are specific for each tRNA species and for each silk gland

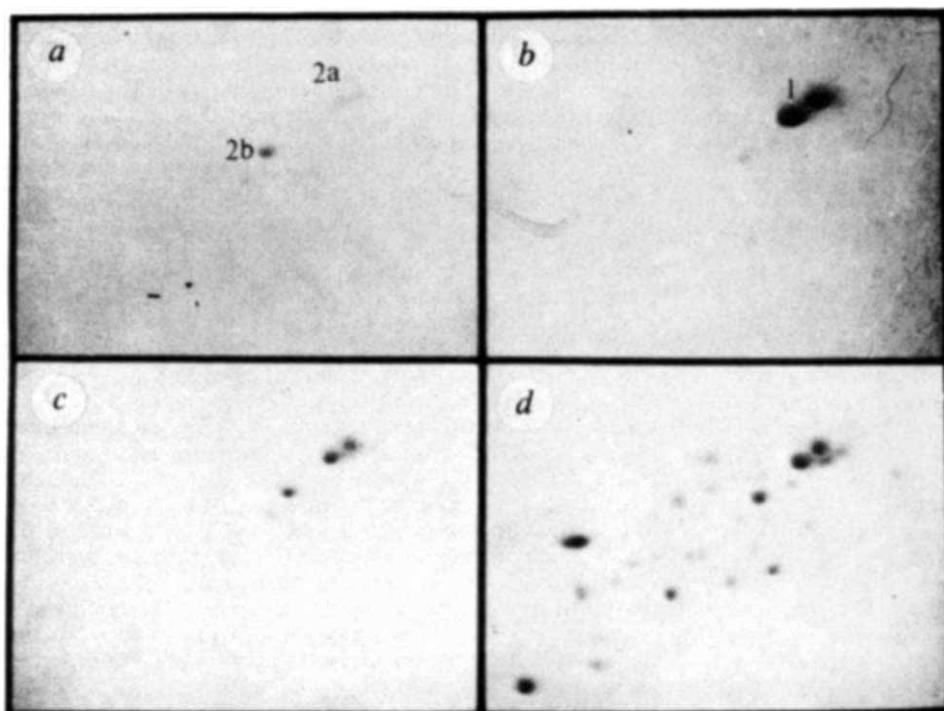


Fig. 2 Identification of tRNA^{Ser} species from the MSG (eighth day from the fifth instar) fractionated on a BD-cellulose column and analysed by two-dimensional polyacrylamide gel electrophoresis. The first dimension was carried out on 10% acrylamide-bisacrylamide (19:1), 7 M urea, Tris-borate buffer 0.089 M, pH 8.3, and run from right to left for 25 h at 12 V cm^{-1} in the cold. The second dimension was done on 20% acrylamide, 4 M urea in the same buffer and run from top to bottom for 48 h at 13 V cm^{-1} . Gels were stained in 0.2% (w/v) methylene blue solution in 2% sodium acetate buffer, pH 5.0, and destained in running water^{13,14}. *a*, tRNA^{Ser} species from the second peak of the BD-cellulose column (fractions 90–150 from Fig. 1b), 0.1 A₂₆₀. *b*, tRNA^{Ser} species from the first peak (fractions 45–80), 0.1 A₂₆₀. *c*, tRNA^{Ser} and tRNA^{Ser} species previously purified from a BD-cellulose column carried out in the absence of Mg^{2+} . *d*, tRNA^{Ser} and tRNA^{Ser} co-electrophoresed with 0.5 A₂₆₀ of total PSG tRNA.

part. It has been shown, for example, that $\text{tRNA}_1^{\text{Gly}}$, $\text{tRNA}_2^{\text{Gly}}$ and $\text{tRNA}_{2a}^{\text{Ala}}$ are the most highly labelled species at all times in the posterior part. Their absolute synthesis rates are at a maximum on days 3 and 4, that is, at the end of the growth phase and the beginning of fibroin secretion. In the middle part however, tRNA^{Asp} , $\text{tRNA}_1^{\text{Gly}}$, $\text{tRNA}_1^{\text{Ser}}$ and $\text{tRNA}_{2a}^{\text{Ser}}$ species have high absolute synthesis rates with a maximum shifted in time, whereas other tRNA species are synthesized at lower rates.

A comparison of biosynthetic rates for the iso- tRNA^{Ser} family is of particular interest. Figure 4 shows that during the fifth instar changes in the relative synthesis rates for $\text{tRNA}_1^{\text{Ser}}$ and $\text{tRNA}_2^{\text{Ser}}$ occur in the two silk gland parts—these changes differ between the two parts. In the PSG, there is an increase in $\text{tRNA}_2^{\text{Ser}}$ synthesis with respect to $\text{tRNA}_1^{\text{Ser}}$ species at the start of fibroin production. However, in the MSG, $\text{tRNA}_1^{\text{Ser}}$ species is preferentially synthesized although the changes are more progressive. We know that these different patterns are essentially due to the highest synthesis rate of $\text{tRNA}_{2a}^{\text{Ser}}$ in the posterior, and $\text{tRNA}_1^{\text{Ser}}$ in the middle part. These individual changes explain

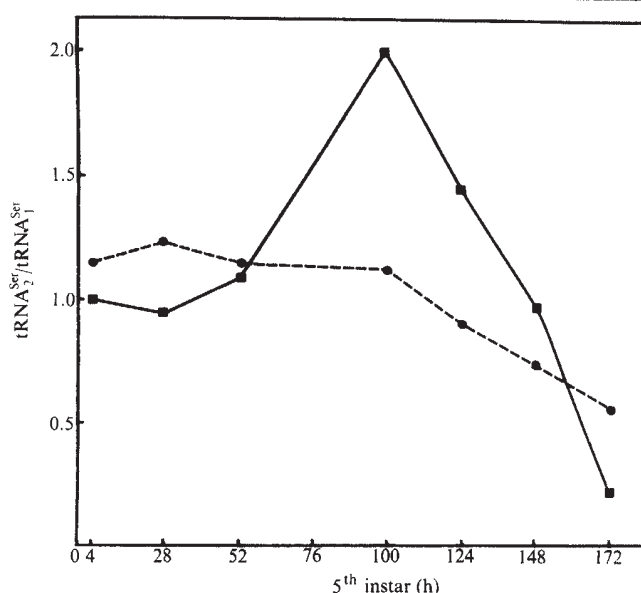


Fig. 4 Differential synthesis rates of $\text{tRNA}_1^{\text{Ser}}$ and $\text{tRNA}_2^{\text{Ser}}$ species in the middle and posterior silk glands of *B. mori* silkworm during the fifth larval instar. Silkworms from days 1 to 8 of the last instar were each injected with 0.5 mCi ^3H -uridine and dissected after 4 h labelling. tRNA from posterior (■) and middle (●) silk glands were subjected to two-dimensional polyacrylamide gel electrophoresis and each tRNA^{Ser} spot quantitated. The ordinate indicates time (h) after the first meal following the last moult.

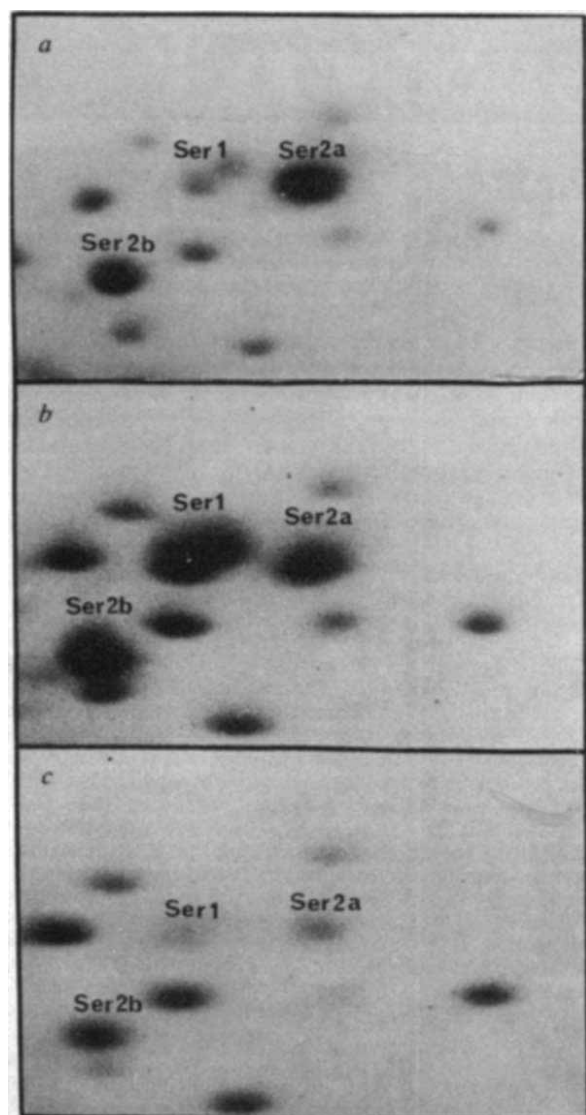


Fig. 3 Differential accumulation of tRNA^{Ser} species on polyacrylamide gel maps from PSG and MSG of *B. mori* silkworm at the end of the fifth larval instar. Two-dimensional electrophoresis was carried out according to Fig. 2 legend. The figures show only the tRNA^{Ser} area corresponding to the upper right part of a complete tRNA electrophoretic map. a, PSG; b, MSG; c, carcass.

the differential accumulations evident in Fig. 1 and shown more precisely in Fig. 3. They seem to be related to the different mRNA populations being decoded.

Of the three possible levels of regulation—preferential degradation of mature tRNA species, selective processing of precursor tRNA, and differential transcription of tRNA genes—we concentrated on the last possibility. The preferential degradation of non-used mature tRNA species can be ruled out, because we have found similar degradation rates for 13 tRNA species, including tRNA^{Ser} species¹³. A selective processing of precursor tRNA, whose primary transcripts would match tRNA gene distribution, cannot be excluded, although preliminary results from silk glands incubated on day 4 for a short time with ^{32}P indicate an enrichment of pre- tRNA^{Ser} in the middle part compared with the posterior part¹³.

Although there is no direct evidence for a transcriptional control of tRNA gene activity according to the needs of the polysomal compartment, the differential appearance and accumulation observed for each iso- tRNA^{Ser} species favour such a mechanism, the most suitable to account for tRNA adaptation to the mRNA codons being translated.

We thank Dr G. Chavancy, A. Fournier and Ms J. Ramsey for critical reading of the manuscript.

Received 14 July 1980; accepted 9 January 1981.

- Daillie, J. *C. r. hebdomadaire. Séances. Acad. Sci., Paris* **261**, 4872–4875 (1965).
- Shigematsu, H. & Takeshita, H. *J. Insect Physiol.* **14**, 1013–1024 (1968).
- Tashiro, Y., Morimoto, T., Matsuura, S. & Nagata, S. *J. Cell Biol.* **38**, 574–588 (1968).
- Fournier, A. *Biochimie* **61**, 283–320 (1979).
- Daillie, J. *Biol. Cell* **29**, 1–6 (1977).
- Garel, J. P., Hentzen, D. & Daillie, J. *FEBS Lett.* **39**, 359–363 (1974).
- Chen, G. S. & Siddiqui, M. A. Q. *Archs Biochem. Biophys.* **161**, 109–117 (1974).
- Majima, R., Kawakami, M. & Shimura, K. *J. Biochem., Tokyo* **78**, 391–400 (1975).
- Garel, J. P. *Nature* **260**, 805–806 (1976).
- Gillam, I. *et al. Biochemistry* **6**, 3043–3056 (1967).
- Garel, J. P., Hentzen, D., Schlegel, M. & Dirheimer, G. *Biochimie* **58**, 1089–1100 (1976).
- Hentzen, D., Garel, J. P. & Keith, G. *Biochem. biophys. Res. Commun.* **71**, 241–248 (1976).
- Chevallier, A. & Garel, J. P. *Biochimie* **61**, 245–262 (1979).
- Garel, J. P., Garber, R. & Siddiqui, M. A. Q. *Biochemistry* **16**, 3618–3624 (1977).
- Rogg, H., Müller, P. & Staehelin, M. *Eur. J. Biochem.* **53**, 115–127 (1975).
- Chavancy, G., Chevallier, A., Fournier, A. & Garel, J. P. *Biochimie* **61**, 71–78 (1979).
- Suzuki, Y. & Brown, D. D. *J. molec. Biol.* **63**, 409–429 (1972).

Human erythrocyte membranes exhibit a cooperative calmodulin-dependent Ca^{2+} -ATPase of high calcium sensitivity

Peter Downes & Robert H. Michell

Department of Biochemistry, University of Birmingham, PO Box 363, Birmingham B15 2TT, UK

It is thought that the ionized Ca^{2+} concentration in the cytosol of healthy erythrocytes is in the range 0.01–0.1 μM (ref. 1) and that this low concentration is maintained by an ATP-driven calmodulin-dependent Ca^{2+} pump in the plasma membrane^{2,3}. The Ca^{2+} -stimulated ATPase which is the enzymatic expression of this pump varies in its calcium sensitivity between different preparations of erythrocyte ghosts, with activation generally occurring in a concentration range between ~1 and 10 μM Ca^{2+} (refs 2–5). This is a higher range of Ca^{2+} concentrations than might be anticipated for activation of a pump that sustains intracellular concentrations of Ca^{2+} below 0.1 μM , and recent reports have suggested activation in some membrane preparations at Ca^{2+} concentrations in the range 0.1–1.0 μM (refs 6, 7). We report here a simple method for preparing human erythrocyte membranes in 2.5 mM HEPES/1 mM EGTA at pH 7.0 (see Fig. 1 legend) in which the activation of the Ca^{2+} -ATPase by Ca^{2+} and intracellular concentrations of calmodulin is highly cooperative and is complete by ~1 μM Ca^{2+} . Unlike other available erythrocyte membrane preparations, the pattern of activation by Ca^{2+} and calmodulin is not complicated by partial resealing of the ghosts during and after isolation. We suggest that this cooperative activation of the Ca^{2+} pump may explain how healthy erythrocytes maintain their normal cytosol Ca^{2+} concentration at a threshold value at or below ~0.1 μM . We also note that several other calmodulin-dependent enzymes display similar cooperative activation kinetics.

We first compared the activation by Ca^{2+} of the calmodulin-dependent ATPase activities of human erythrocyte ghosts that had been prepared in various media, but which were all assayed in conditions of ionic strength and of ATP and Mg^{2+} concentration relatively similar to those in the cell. We observed that membranes isolated in 2.5 mM HEPES/1 mM EGTA at pH 7.0 exhibited the behaviour shown in Fig. 1a: activation occurred at lower Ca^{2+} concentrations than with other ghost preparations, and full activation occurred over a narrow range of Ca^{2+} concentrations. A rise from 10 to 90% of full activity required only a four- to fivefold increase in Ca^{2+} concentration, indicating that Ca^{2+} activated the ATPase in a highly cooperative manner (Hill coefficient 3.0 ± 0.4 , $n = 5$). In the presence of a high concentration of calmodulin, substantial activity was expressed at least down to 0.1 μM (Fig. 1a). As the concentration of calmodulin was reduced, however, the proportion of the total activity expressed at micromolar concentrations of Ca^{2+} fell (Fig. 1a, b). These experimental results are very similar to those reported by Klinger *et al.*⁷ but these authors interpreted their data in terms of classical Michaelis–Menten kinetics and thus, we feel, overlooked the potential importance of their observations.

If these results reflect the pattern of activation of the Ca^{2+} pump in the intact erythrocyte, into which Ca^{2+} leaks only very slowly, the ionized Ca^{2+} concentration within the cell would automatically be maintained below 0.1 μM , that is, at the expected level for a healthy erythrocyte. That this pattern may indeed be the correct one, rather than an experimental artefact, was indicated by comparative experiments (not shown) in which the Ca^{2+} dependence of the Ca^{2+} -ATPase was assessed in several membrane preparations, using both unmodified samples and samples treated with 0.1 mg per ml saponin^{7,8}; the preparations tested were the new HEPES/EGTA ghosts, whole

erythrocytes, ghosts prepared in 20 mM Tris/1 mM EGTA and ghosts prepared in imidazole/histidine/EGTA mixtures⁷. In the absence of saponin, the three latter preparations showed either no ATPase (cells) or activation that was much less sensitive to Ca^{2+} and calmodulin than in the HEPES/EGTA preparation, but when saponin was present (to ensure membrane permeability) they all behaved like the saponin-free HEPES/EGTA ghosts illustrated in Fig. 1. We therefore believe that much of the reported variability in Ca^{2+} and calmodulin sensitivity of previously described ghost preparations arose simply because they were partially (and variably) resealed, so that access of added Ca^{2+} , ATP or calmodulin to the ATPase was restricted.

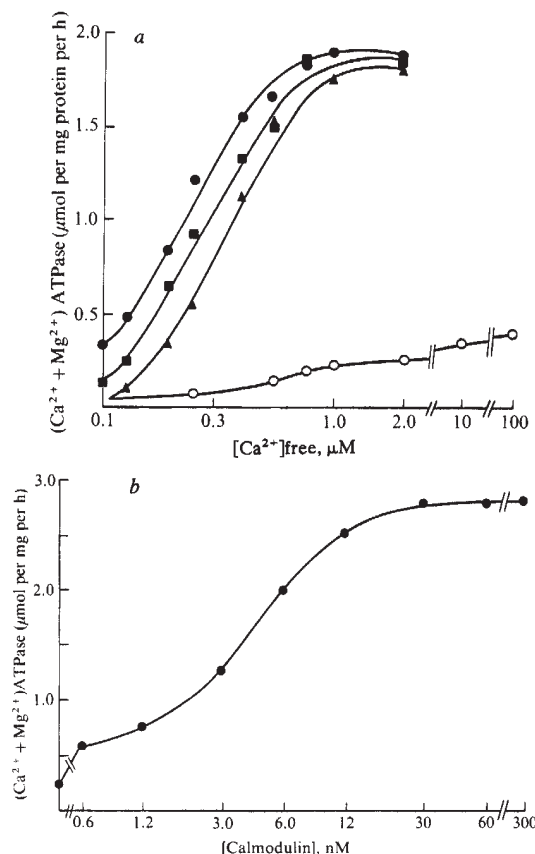


Fig. 1 Activation of $(\text{Ca}^{2+} + \text{Mg}^{2+})\text{ATPase}$ by *a*, Ca^{2+} in the absence (○) or presence of bovine brain calmodulin (▲, 56 nM; ■, 280 nM; ●, 5,600 nM), and *b*, calmodulin in the presence of a maximally stimulating concentration of Ca^{2+} (2 μM). Human blood was obtained from the Regional Blood Transfusion Laboratories within 7 days of collection. Erythrocytes were sedimented and washed three times in 0.15 M NaCl/1.5 mM HEPES, pH 7.2, at 4 °C to remove plasma and leukocytes. Cells were lysed in 10 vol of 2.5 mM HEPES/1 mM EGTA, pH 7.0. The membranes were sedimented at 15,000g for 15 min and then washed five times in the same medium. The isolated membranes were stored in this medium at 4 °C for up to 24 h before use. $(\text{Ca}^{2+} + \text{Mg}^{2+})\text{ATPase}$ activity was assayed for 30 min at 30 °C in 1 ml of 30 mM HEPES, 30 mM KCl, 80 mM NaCl, 4 mM MgCl_2 , 3 mM ATP, 0.1 mM ouabain, 1 mM EGTA, to which various concentrations of Ca^{2+} were added before adjustment of the pH to 7.0. Calmodulin was isolated from bovine brain¹⁶. Assays contained 0.15–0.2 mg of membrane protein. The ionized Ca^{2+} concentrations were calculated using $k_{\text{Ca}^{2+}\text{-EGTA}}$ of 11.0, as recommended by Owen¹⁷. Reactions were stopped by addition of cold trichloroacetic acid (final concentration 10% w/v) and P_i in the acid-soluble fraction measured¹⁸. Ca^{2+} -ATPase activity was calculated as total ATPase activity minus the activity expressed in the presence of 1 mM EGTA without added Ca^{2+} . The results shown in *a* and *b* are each taken from duplicate incubations in a single experiment, and are typical of results obtained in several experiments of fundamentally similar design. Results of the type shown were consistently obtained in many preparations of ghosts isolated in 2.5 mM HEPES, 1 mM EGTA, pH 7.0, but not in ghosts isolated in various other conditions.

Both we (Fig. 1) and others (for example, ref. 7) have noted that erythrocyte Ca^{2+} -ATPase can express its full activity at concentrations of calmodulin far lower than those present in the cell ($\sim 5 \mu\text{M}$)^{2,7}. Our results (closed symbols in Fig. 1a) and those of Klinger *et al.*⁷ also demonstrate that the calmodulin-dependent activation of the ATPase by Ca^{2+} retains its cooperativity at reduced calmodulin concentrations. However, as the calmodulin concentration is reduced, the apparent affinity for Ca^{2+} declines, suggesting that a function of the high, apparently 'excessive', calmodulin concentration that prevails within cells may be to maximize the sensitivity to Ca^{2+} of an already sensitive Ca^{2+} pump. Further studies using concentrations of Ca^{2+} and/or calmodulin that were inadequate for full activation gave complex results (not shown) that seem to rule out any explanation of activation involving simply the generation of a certain concentration of a particular Ca^{2+} -calmodulin complex as suggested by Cheung⁹; the relationship of these results to the mechanism of cooperative activation will be considered elsewhere.

At $\sim 1.0 \mu\text{M}$ Ca^{2+} , at which full activation of the ATPase was seen in the presence of calmodulin, the 'basal' (that is, calmodulin-independent) Ca^{2+} -ATPase activity of these ghosts was no more than 5% of their maximum activity, but this could be raised, in a relatively non-cooperative manner (open symbols in Fig. 1a), to $\sim 20\%$ of maximum activity at a Ca^{2+} concentration of $\sim 100 \mu\text{M}$. This activity persisted through extensive and varied washing procedures, as did a residual complement of $62 \pm 15 \text{ ng}$ ($n = 3$) of membrane-associated calmodulin per mg of membrane protein (as assayed by the activation of skeletal muscle myosin light chain kinase by heat-denatured erythrocyte ghosts; H. Dolbear and P.D., unpublished results). It therefore seems that soluble calmodulin must be available if the activation of the ATPase by Ca^{2+} is to show its normal cooperativity, but it remains possible that control by Ca^{2+} of the lower-affinity, 'calmodulin-independent' basal activity may still be mediated, in a different manner, by residual calmodulin that is tightly bound in the membrane preparations, as has been suggested previously^{7,10}. If this is correct, this residual calmodulin might be as tightly bound to the membrane ATPase as the calmodulin that is bound into the subunit structure of phosphorylase kinase.

Recent work has suggested that some Ca^{2+} -mobilizing hormones and neurotransmitters evoke intracellular responses by causing relatively small (for example, fivefold) changes in the cytosolic ionized Ca^{2+} concentration (for example, ref. 11). Such small changes in Ca^{2+} concentration should therefore be able to control the key intracellular enzymes responsible for these responses. At least two calmodulin-dependent Ca^{2+} -activated enzymes other than the Ca^{2+} -ATPase (myosin light chain kinase^{12,13} and cyclic nucleotide phosphodiesterase¹⁴) show highly cooperative activation by low concentrations of Ca^{2+} . Recent data suggest that the same may also apply to glycogen phosphorylase kinase when activated through its endogenous calmodulin subunit but not when activated by exogenous troponin I; this cooperativity was not, however, noted in the original paper¹⁵. It seems possible that such cooperativity may be an essential and general characteristic for the functioning of many enzymes that must respond to small changes in cytosol Ca^{2+} concentration. It will be interesting to determine whether the same type of mechanism underlies the cooperative activation by Ca^{2+} of all these calmodulin-dependent enzymes.

We thank the MRC for the award of a research studentship to P.D.

Received 4 November 1980; accepted 5 February 1981.

- Schatzmann, H. J. *Curr. Topics Membranes Trans.* **6**, 125-168 (1975).
- Jarrett, H. W. & Penniston, J. T. *Biochem. biophys. Res. Commun.* **77**, 1210-1216 (1977).
- Gopinath, R. M. & Vincenzi, F. F. *Biochem. biophys. Res. Commun.* **77**, 1203-1209 (1977).
- Scharff, O. & Foder, B. *Biochim. biophys. Acta* **509**, 67-77 (1978).
- Hanahan, D. J., Taverna, R. D., Flynn, D. D. & Ekholm, J. E. *Biochem. biophys. Res. Commun.* **84**, 1009-1015 (1978).
- Muallem, S. & Karlisch, S. J. D. *Biochim. biophys. Acta* **597**, 631-636 (1980).
- Klinger, R., Wetzker, R., Fleischer, I. & Frunder, H. *Cell Calcium* **1**, 229-240 (1980).
- Hanahan, D. J. & Ekholm, J. E. *Archs Biochem. Biophys.* **187**, 170-179 (1978).
- Cheung, W. Y. *Science* **207**, 19-27 (1980).
- Lynch, T. J. & Cheung, W. Y. *Archs Biochem. Biophys.* **194**, 165-170 (1979).

- Murphy, E., Coll, K., Rich, T. L. & Williamson, J. R. *J. biol. Chem.* **255**, 6600-6608 (1980).
- Nairn, A. C. & Perry, S. V. *Biochem. J.* **179**, 89-97 (1979).
- Stull, J. T. *Adv. Cyclic Nucleotide Res.* **13**, 39-93 (1980).
- Dedman, J. R., Potter, J., Jackson, R. L., Johnson, J. D. & Means, A. R. *J. biol. Chem.* **252**, 8415-8422 (1977).
- Cohen, P. *Eur. J. Biochem.* **111**, 563-574 (1980).
- Grand, R. J. A., Perry, S. V. & Weeks, R. A. *Biochem. J.* **177**, 521-529 (1979).
- Owen, J. D. *Biochim. biophys. Acta* **451**, 321-325 (1976).
- Baginski, E. S., Zak, B. & Foa, P. P. *Clin. Chem.* **13**, 326-332 (1967).

Occluded bound calcium on the phosphorylated sarcoplasmic transport ATPase

H. Takisawa & M. Makinose

Max-Planck-Institut für medizinische Forschung, Heidelberg, Abteilung Physiologie, 6900 Heidelberg, Jahnstrasse 29, FRG

The Ca^{2+} + Mg^{2+} -activated ATPase of the sarcoplasmic reticulum is responsible for the active Ca^{2+} transport of this membrane system¹, the key feature of which is the formation of an energy-rich phosphorylated transport enzyme (EP) and its conversion². To understand the Ca^{2+} -transport mechanism, it is essential to clarify the behaviour of this intermediate in relation to such ligands as ATP, ADP, Mg^{2+} and, particularly, Ca^{2+} . Recent kinetic studies on the phosphate turnover of this system suggested a relatively slow rate of Ca^{2+} dissociation from the phosphorylated enzyme^{3,4}, which possibly indicated Ca^{2+} binding in some occluded form with the intermediate. Here we report direct measurements of the binding and release of Ca^{2+} during phosphorylation of the sarcoplasmic transport enzyme. The results indicate an occlusion of the Ca^{2+} binding, accompanying an initial configurational change of the enzyme induced by the energy-rich phosphoryl transfer.

Figure 1 shows Scatchard plots of Ca^{2+} binding by the vesicular and solubilized sarcoplasmic transport ATPase in the absence of ATP. No essential difference could be observed between assays using 1 and 5 mM MgCl_2 . The curves for both preparations show a break. In the case of vesicular ATPase, the calcium affinities of the preparation, calculated from the slopes of the curves, were $k_{\text{diss}} \geq 1 \mu\text{M}$ and $\approx 10 \mu\text{M}$, which suggests negatively cooperating sites for Ca^{2+} binding^{5,6}. The maximal capacity for Ca^{2+} binding, obtained by extrapolation of the curves in Fig. 1 to the ordinate intercept, corresponds to 2.2-2.3 mol Ca^{2+} per mol of the ATPase phosphorylatable site in the vesicular preparation. Numerous kinetic analyses of the transient and/or steady state activity of the native or the partially purified vesicular sarcoplasmic ATPase have shown that one molecule of the transport ATPase is activated by the binding of two Ca^{2+} (ref. 7). The maximal Ca^{2+} binding capacity obtained from Fig. 1 includes, therefore, 0.2-0.3 mol of Ca^{2+} binding capacity per mol phosphorylatable site in excess. This could be due to the nonspecific Ca^{2+} binding capability of the preparation, for example, due to contamination of the Ca^{2+} -binding protein or of denatured ATPase protein.

All the ATPase preparations formed, with ATP, a maximum of 4-5 nmol phosphoenzyme per mg protein. Based on a molecular weight of 115,000 for the sarcoplasmic ATPase⁸, this indicates that about half of the vesicular or solubilized ATPase preparations consist of enzymatically inert proteins—that is, mainly denatured ATPase protein.

However, the solubilized ATPase protein shows a maximal Ca^{2+} binding capacity of ~ 1.4 mol per mol of phosphorylatable site. As both the ATPase preparations show the same specific activity, it is reasonable to assume that they both contain the same amount of inert protein. Thus, the subtraction of nonspecific Ca^{2+} binding capacity found in the vesicular preparation gave a maximal Ca^{2+} specific binding capacity of the solubilized ATPase which was almost half that of the former preparation. A break in the Scatchard plots of solubilized

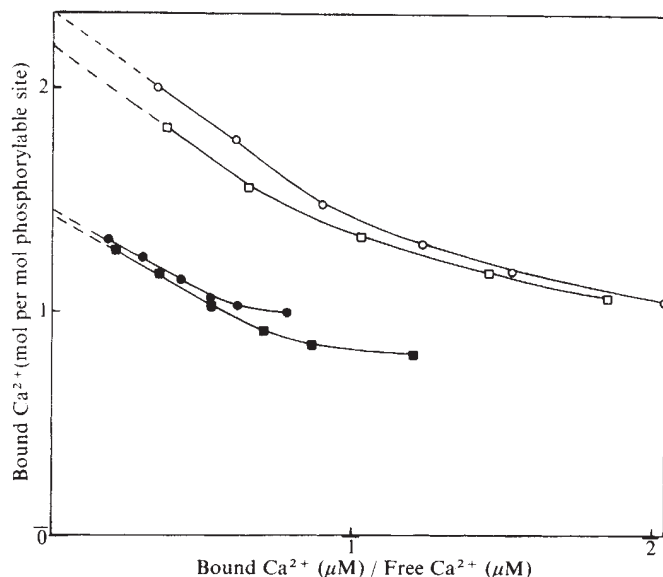


Fig. 1 Scatchard plots of Ca^{2+} binding by the vesicular and solubilized Ca^{2+} -transport ATPase. Binding of Ca^{2+} to the ATPase was determined by the flow-dialysis method of Colowick and Womack¹⁰ in solutions containing 0.4 M KCl, 0.3 M sucrose, 1 (○, ●) or 5 mM (□, ■) MgCl_2 and 50 mM Tris-TES (trimethylamino ethanesulphonic acid), pH 8, at 5 °C. The concentration of the enzymes was adjusted to 3 mg ml⁻¹ and, in the case of solubilized ATPase (●, ■), 3.5 mg ml⁻¹ deoxycholate (DOC) was included in the assay. Vesicular ATPase (partially purified sarcoplasmic reticulum vesicles: ○, □) was prepared by the method of Meissner *et al.*¹¹. To obtain the solubilized ATPase, 3.5 mg ml⁻¹ of the sarcoplasmic protein was solubilized with DOC (3.5 mg ml⁻¹) in the presence of 0.4 M KCl, 0.3 M sucrose and 50 mM Tris-TES, pH 8, at 5 °C¹². After pelleting of the non-solubilized proteins (105,000g for 50 min), the supernatant contained ~3 mg of protein and 3.5 mg of DOC per ml. About 6 nmol Ca^{2+} per mg protein (obtained by atomic absorption spectrometry, Perkin-Elmer Model 4000) constituted Ca^{2+} contamination in both enzyme preparations. The measurement of Ca^{2+} binding was started by the addition of 4 nmol (3.2 μCi) $^{45}\text{Ca}^{2+}$ to 1 ml of the enzyme solution. The Ca^{2+} concentration was increased stepwise up to 2 mM by the addition of 'cold' CaCl_2 . The ordinate is the amount of Ca^{2+} bound per mol of phosphorylatable site of the preparation used. The maximum amount of phosphoprotein was estimated in an assay containing 3 mM CaCl_2 , 1 mM MgCl_2 and 100 μM ATP, as 4–5 nmol EP per mg protein for both the vesicular and solubilized ATPase.

ATPase (Fig. 1) is presumably due to the nonspecific bound Ca^{2+} capacity. Apparently, the solubilization reduces the affinity of one of the two sites of the enzyme for Ca^{2+} .

This raises the question of whether the solubilized sarcoplasmic ATPase requires one or two Ca^{2+} for its activation. Because the rate of the EP formation of the solubilized ATPase is much slower than that of the vesicular ATPase (data not shown), the time course of EP formation can be easily followed. Figure 2 shows the change of EP level, extra Ca^{2+} binding and P_i liberation during transient EP formation. It is evident that the extra binding of Ca^{2+} was accompanied by EP formation and the amount of additional Ca^{2+} bound to the ATPase was almost equal to the amount of EP formed. No correlation was found between the time course of additional Ca^{2+} binding and P_i liberation. Apparently, two Ca^{2+} are also required for the activation of the solubilized sarcoplasmic ATPase. The reaction steps for the activation of the solubilized ATPase can thus be formulated:

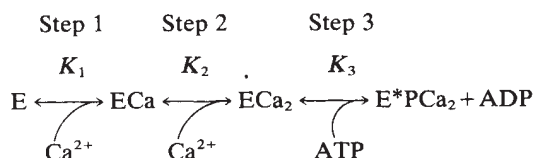


Table 1 Retention of bound calcium on the phosphorylated sarcoplasmic transport enzyme

ATPase preparation	Addition	Effluent contents		
		Calcium* (nmol)	Phosphorylatable site (nmol)	Mol bound Ca^{2+} per mol phosphorylatable site
Vesicular	None	0.180 ± 0.005	0.244 ± 0.004	0.739 ± 0.010
	None	0.099 ± 0.005	0.306 ± 0.004	0.324 ± 0.015
Solubilized	ATP (100 μM)	0.371 ± 0.011	0.259 ± 0.019	1.437 ± 0.076
	AMPPCP (200 μM)	0.091 ± 0.007	0.281 ± 0.007	0.322 ± 0.024

A small Sephadex G-50 column (4.7 × 57 mm) prepared according to Penefsky⁹ was equilibrated with a solution containing 0.4 M KCl, 0.3 M sucrose and 50 mM Tris-TES, pH 8. The enzymes (3 mg protein ml⁻¹), were incubated with $^{45}\text{CaCl}_2$ (1 mM) in the presence or absence of ATP (100 μM) or AMPPCP (200 μM) in a solution containing 0.4 M KCl, 0.3 M sucrose, 80 μM MgCl_2 and 50 mM Tris-TES, pH 8 at 0 °C for 2 min. An aliquot of the assay (50 μl) was transferred to the top of the column and centrifuged immediately (150g for 2 min at 5 °C) and the effluent calcium and protein content were measured. The enzyme preparations showed the maximal phosphorylatable site number of 5.1 nmol per mg protein using the method described in Fig. 1 legend, with no significant difference between the vesicular and the solubilized enzyme. Values are the means of four determinations ± s.e.m.

* The calcium content of the effluent in the control experiment containing no protein was <0.002 nmol.

The solubilizing procedure affects the ATPase by reducing its affinity for the second Ca^{2+} (K_2), so that, at the concentration of Ca^{2+} which is usually sufficient to activate the transport enzyme (for example, 90 μM as in Fig. 2), one solubilized ATPase molecule binds about one Ca^{2+} . If ATP is added, step 3 of the reaction is activated and the enzyme is accumulated in the form of E^*PCa_2 , which retains two Ca^{2+} per molecule. The binding state of the Ca^{2+} with this enzyme species should be very different from that of the unphosphorylated enzyme species.

In the experiments shown in Table 1, the ATPase preparations, after incubation in appropriate conditions, were centrifuged through a small Sephadex column to rapidly remove the incubation medium (centrifuge column method⁹). As the enzyme is exposed to the ligand free-solution in the column

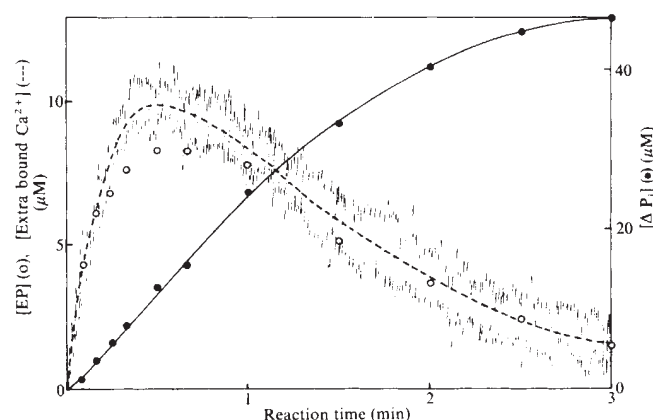


Fig. 2 Extra Ca^{2+} binding to the solubilized sarcoplasmic ATPase during EP formation. The reaction was started by adding 50 μM $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ to the assay containing 3 mg ml⁻¹ solubilized ATPase protein, 3.5 mg ml⁻¹ DOC, 0.4 M KCl, 0.3 M sucrose, 90 μM CaCl_2 , 80 μM MgCl_2 and 50 mM Tris-TES pH 8.0, at 5 °C and terminated after the appropriate time by addition of 4% trichloroacetic acid containing 0.1 mM ATP and 0.2 mM P_i . EP formation and P_i liberation were measured as described elsewhere^{13,14}. Change in the concentration of Ca^{2+} was measured in the same assay containing 100 μM Murexide¹⁵. The reaction was started with quick addition (within 1 s) of 50 μM cold ATP and the absorbance change of the dye at 507/540 nm was registered on a Aminco DW-2 spectrophotometer. The addition of ATP caused an immediate change in absorbance corresponding to the decrease in Ca^{2+} concentration of ~2 μM due to formation of Ca -ATP complex. This deviation was subtracted from the original trace.

during the separation time, the bound ligands, particularly the rapidly exchangeable ligands, are partially removed from the enzyme protein. When the vesicular ATPase was preincubated in solution containing 1 mM Ca^{2+} , 0.74 mol of Ca^{2+} was bound per mol active enzyme after centrifugation instead of the expected 2 mol. In the case of solubilized ATPase, 0.32 mol of Ca^{2+} was bound per mol of ATPase instead of the expected 1 mol. Apparently, about two thirds of the rapidly exchangeable bound Ca^{2+} is removed from the ATPase protein in this procedure. When the solubilized ATPase was incubated with 100 μM ATP and 1 mM Ca^{2+} , 1.44 mol of Ca^{2+} were bound per mol active enzyme in the effluent. In a parallel assay with 'hot' ATP and 'cold' Ca^{2+} , in the effluent, 32% of the phosphorylatable sites of the enzyme were unphosphorylated and 68% were phosphorylated. Therefore, 1 mol of the active ATPase in these effluents retains 0.1 mol (32% of 0.32 mol) Ca^{2+} , which represents the unphosphorylated state and the remaining 1.34

mol Ca^{2+} represents the phosphorylated state (68% of 1 mol). Thus, even after column centrifugation, 2 mol of Ca^{2+} are retained by 1 mol phosphorylated enzyme. Addition of 200 μM AMPPCP (non-hydrolysable ATP analogue) had no effect on the Ca^{2+} -binding of the solubilized ATPase.

The condition for EP formation in which bound Ca^{2+} cannot be removed is low Mg^{2+} concentration (80 μM). When the experiment in Fig. 2 was done using 5 mM Mg^{2+} , addition of ATP caused a transient increase in Ca^{2+} concentration in the assay, as observed previously by Ikemoto⁵. A high Mg^{2+} concentration presumably effects the further conversion of the phosphoprotein, resulting in Ca^{2+} liberation and dephosphorylation. The occlusion of the bound calcium on the transport ATPase molecule is coupled with an initial configurational change of the enzyme molecule caused by phosphorylation of the enzyme protein, but not by formation of the substrate-enzyme complex.

Received 28 July 1980; accepted 2 January 1981.

- Hasselbach, W. & Makinose, M. *Biochem. Z.* **339**, 99–111 (1963).
- Tonomura, Y. *Muscle Proteins, Muscle Contraction and Cation Transport* (University of Tokyo and University Park Press, Tokyo and Baltimore, 1972).
- Takakuwa, Y. & Kanazawa, T. *Biochem. biophys. Res. Commun.* **88**, 1209–1216 (1979).
- Takisawa, H. & Tonomura, Y. *J. Biochem., Tokyo* **86**, 425–441 (1979).
- Ikemoto, N. *J. Biol. Chem.* **250**, 7219–7224 (1975).
- Yates, D. W. & Duance, V. C. *Biochem. J.* **159**, 719–728 (1976).
- Yamamoto, T., Takisawa, H. & Tonomura, Y. in *Current Topics in Bioenergetics* Vol. 9 (ed. Sanadi, D. R.) (Academic, New York, 1979).
- Le Maire, M., Jørgensen, K. E., Røigaard-Petersen, H. & Møller, J. V. *Biochemistry* **15**, 5805–5812 (1976).
- Penefsky, H. S. *J. Biol. Chem.* **252**, 2891–2899 (1977).
- Colowick, S. P. & Womack, F. C. *J. Biol. Chem.* **244**, 774–777 (1969).
- Meissner, G., Conner, G. E. & Fleischer, S. *Biochim. biophys. Acta* **298**, 246–269 (1973).
- Jørgensen, K. E., Lind, K. E., Røigaard-Petersen, H. & Møller, J. V. *Biochem. J.* **169**, 489–498 (1978).
- Makinose, M. *Eur. J. Biochem.* **10**, 74–82 (1969).
- Takisawa, H. & Tonomura, Y. *J. Biochem., Tokyo* **83**, 1275–1284 (1978).
- Scarpa, A. *Meth. Enzym.* **24**, 343–351 (1972).

Direct observation by NMR of two coexisting conformations of an enzyme–ligand complex in solution

A. Gronenborn, B. Birdsall, E. I. Hyde, G. C. K. Roberts*, J. Feeney & A. S. V. Burgen

Division of Molecular Pharmacology, National Institute for Medical Research, London NW7 1AA, UK

Dihydrofolate reductase is the site of action of the 'anti-folate' drugs, such as methotrexate, pyrimethamine and trimethoprim, which act as competitive inhibitors of the enzyme^{1–3}. These inhibitors bind to the *Lactobacillus casei* enzyme cooperatively with the coenzyme NADP(H), which can increase inhibitor binding by as much as 700-fold⁴. The existence of conformational changes accompanying ligand binding and involved in inhibitor–coenzyme cooperativity has been inferred from NMR experiments^{4–9}. It seems likely that these conformational changes are important in determining the specificity of the enzyme for its ligands^{4,10,11}. We now report the direct observation by NMR of two distinct conformations of the enzyme–trimethoprim–NADP⁺ ternary complex in solution.

Figure 1 compares the histidine C^ε–H resonances of this complex with those of the enzyme–methotrexate–NADP⁺ complex. The resonances of all seven histidine residues¹², labelled A–G, are clearly resolved in these conditions. Their positions are closely similar in the two complexes (for a detailed comparison, see ref. 13), with the exception of resonance F. This has been assigned to His 28 (ref. 14), which is seen in the crystal structure¹⁵ of the enzyme–methotrexate–NADPH complex to interact with the γ -carboxyl of the glutamate moiety of methotrexate. As there is no corresponding group in trimethoprim, the pK of His 28 is ~ 1.3 units lower in complexes with trimethoprim than in those with methotrexate^{13,16}.

The most striking difference between the two spectra in Fig. 1, however, is that in the spectrum of the trimethoprim complex resonances E (tentatively assigned to His 18 (refs 13, 14)) and F are split into two peaks, each of intensity corresponding to approximately half a proton (compare resonances A, B and D).

The separation between these two components of resonances E and F, measured in hertz, is directly proportional to magnetic field strength and therefore represents a chemical shift difference. The spectra thus show that the enzyme–trimethoprim–NADP⁺ complex exists in two forms which differ in the environments of two histidine residues, His 28 and, probably, His 18. The approximately equal intensity of the two components of the histidine resonances indicates that these two forms are present in approximately equal amounts.

If these two forms interconvert with one another, they must represent different conformational states of the complex. The observations of two distinct signals separated by only ~ 15 Hz (at 270 MHz and 5 °C) implies that such an interconversion must be slow. We have demonstrated that interconversion occurs and measured its rate by determining the temperature dependence of the line shape of histidine resonances E and F over the range 0–50 °C at both 270 and 470 MHz. The results at 470 MHz are shown in Fig. 2.

As the temperature is increased, the two components of each histidine resonance broaden, and then coalesce into a single line which sharpens as the temperature is increased further. This behaviour is that expected for a progressively increasing rate of interconversion between the two states represented by the two components of the absorption band. (The progressive downfield shift of the resonances with increasing temperature is due to the temperature dependence of the pK values of the histidine residues.) These changes in line shape, and particularly the broadening of the resonances with increasing temperature at low temperature, cannot be accounted for by any model in which there is no interconversion between the two states, thus eliminating the possibility that they correspond to two hitherto undetected isoenzymes of the kind observed in *Escherichia coli* RT500 (ref. 17).

The observed line shapes can be satisfactorily simulated, as indicated in Fig. 2, on the basis of a model involving interconversion between just two states of the protein. These simulations allow us to estimate the rate of interconversion at each temperature, taking the chemical shift difference between the two states as that measured at 5 °C (0.044 p.p.m. for resonance F, 0.063 p.p.m. for resonance E). The estimated rates range from 2 s^{–1} at 26 °C to 80 s^{–1} at 51 °C. Four independent estimates of the rate can be made at each temperature from simulation of resonance E and resonance F each at 270 and 470 MHz. These four estimates always agreed within $\pm 15\%$,

* To whom correspondence should be addressed.

indicating that the two-state model of the interconversion process is adequate. The fact that the two components of resonance F coalesce at a lower temperature than those of resonance E can be simply explained by the smaller chemical shift difference between the two states in the former case. The activation parameters for the interconversion, calculated from the temperature dependence of the estimated rate constants, are: $\Delta H^\ddagger = 25(\pm 3)$ kcal mol⁻¹, $\Delta S^\ddagger = 29(\pm 8)$ cal K⁻¹ mol⁻¹. There is no detectable change in the relative proportions of the two states over the temperature range studied.

Note that the splitting of the two histidine resonances is not observed in either the enzyme-trimethoprim or the enzyme-NADP⁺ binary complexes. The observation of two coexisting conformations of the ternary complex but not of either binary complex is clear evidence for the existence of conformational changes specific to the ternary complex. These may be involved in the observed cooperativity in binding between NADP⁺ and trimethoprim.

Examination of the three-dimensional structure of the enzyme-methotrexate-NADPH complex¹⁵ shows that the backbone carbonyl group of His 18 forms a hydrogen bond to the 3'-hydroxyl of the nicotinamide ribose of the bound coenzyme. His 28, on the other hand, is farther both from the bound coenzyme and from trimethoprim (bound in the conformation deduced from ¹H NMR data¹⁸). We have recently shown¹⁹ that the interaction of the nicotinamide ring of the coenzyme with the enzyme and the conformation of the bound coenzyme both differ between the two conformations of the ternary complex. The two environments of His 18 may thus reflect the two modes of interaction of the coenzyme. The effect on His 28, however, indicates that the conformational differences between the two states of the complex extend beyond the residues in direct contact with the bound ligands.

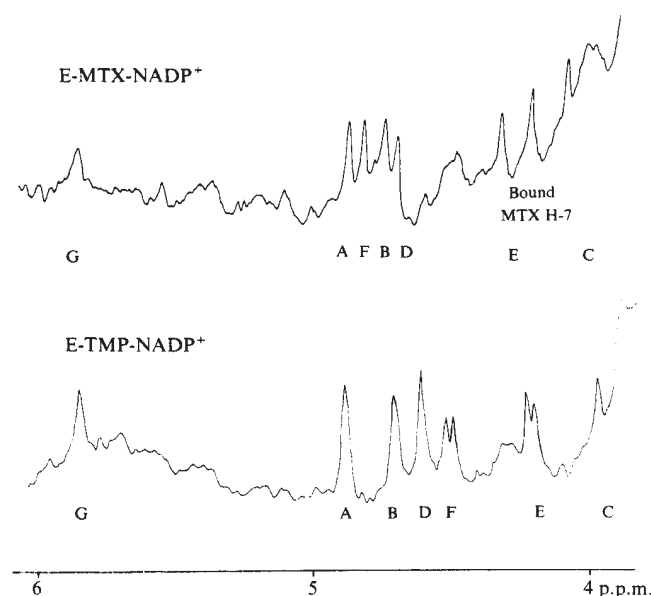


Fig. 1 Histidine C^ε₁-proton resonance region of the ¹H NMR spectrum of *L. casei* dihydrofolate reductase in its complexes with methotrexate (MTX) and NADP⁺ (top) or trimethoprim (TMP) and NADP⁺ (bottom). The resonances of the seven histidine residues are labelled A–G. The spectra were obtained at 270 MHz using a Bruker WH270 spectrometer in the Fourier transform mode. The samples consisted of ~1 mM enzyme in 0.35 ml² H₂O containing 50 mM potassium phosphate, pH (meter reading) 6.5, 500 mM KCl. Approximately 2,000 transients obtained using quadrature detection, with 8,192 data points for a spectral width of 4.2 kHz, were averaged for each spectrum. Before Fourier transformation, the free induction decay was multiplied by an exponential equivalent to a line broadening of 2 Hz. Chemical shifts are expressed relative to internal (1 mM) dioxan (3.71 p.p.m. downfield from dimethylsilapentane-5-sulphonate).

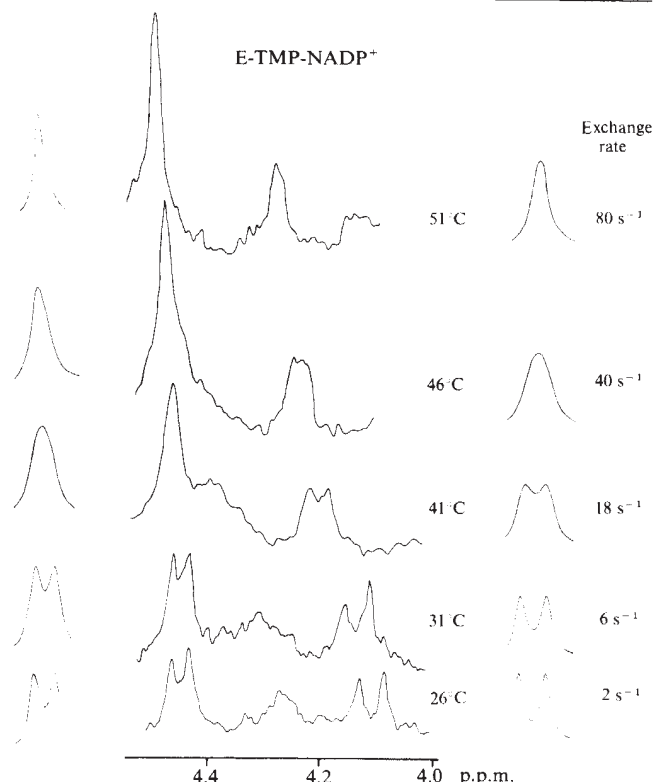


Fig. 2 The histidine C^ε₁-proton resonances E (on the right) and F of the dihydrofolate reductase-trimethoprim-NADP⁺ complex as a function of temperature. The line shapes simulated using the equation for a two-site exchange process, with the chemical shift difference measured at 5 °C and the rate constants shown (chosen for the best visual fit to the experimental spectra), are shown at either side. The experimental spectra were obtained as described in Fig. 1 legend, but using the 470-MHz spectrometer of the Oxford Enzyme Group.

The coexistence of two or more conformational states of an enzyme complex has often been postulated, or inferred from indirect evidence, in this and other systems. In the present case, however, we are able to observe two coexisting conformational states directly in solution, measure their rate of interconversion and begin to characterize the structural differences between them.

We thank Gill Ostler and John McCormick for technical assistance and the Oxford Enzyme Group, particularly Iain Campbell and Peter Stiles, for the use of their 470-MHz NMR spectrometer. E.I.H. is supported by a MRC Research studentship.

Received 1 September 1980; accepted 5 February 1981.

1. Blakley, R.R. L. *The Biochemistry of Folic Acid and Related Pteridines* (North-Holland, Amsterdam, 1969).
2. Kisliuk, R. L. & Brown, G. M. (eds) *Chemistry and Biology of Pteridines* (Elsevier, Amsterdam, 1979).
3. Burchall, J. J. & Hitchings, G. H. *Molec. Pharmac.* **1**, 126–136 (1965).
4. Birdsall, B., Burgen, A.S.V. & Roberts, G. C. K. *Biochemistry* **19**, 3723–3731 (1980).
5. Hyde, E. I., Birdsall, B., Roberts, G. C. K., Feeney, J. & Burgen, A. S. V. *Biochemistry* **19**, 3738–3746 (1980).
6. Hyde, E. I., Birdsall, B., Roberts, G. C. K., Feeney, J. & Burgen, A. S. V. *Biochemistry* **19**, 3746–3754 (1980).
7. Feeney, J. *et al. Proc. R. Soc. B* **196**, 267–290 (1977).
8. Roberts, G. C. K. *et al. in NMR in Biology* (eds Dwek, R. A., Campbell, I. D., Richards, R. E. & Williams, R. J. P.) 95 (Academic, London, 1977).
9. Feeney, J. *et al. Biochemistry* **19**, 2466–2472 (1980).
10. Roberts, G. C. K. *Ciba Fdn Symp.* **60**, 89–103 (1978).
11. Birdsall, B., Burgen, A. S. V., de Miranda, J. P. R. & Roberts, G. C. K. *Biochemistry* **17**, 2102–2110 (1978).
12. Bitar, K. G. *et al. FEBS Lett.* **80**, 119–122 (1977).
13. Gronenborn, A. *et al. Biochemistry* (in the press).
14. Wyeth, P. *et al. Biochemistry* **19**, 2608–2615 (1980).
15. Matthews, D. A. *et al. J. biol. Chem.* **253**, 6946–6954 (1978).
16. Birdsall, B., Griffiths, D. V., Roberts, G. C. K., Feeney, J. & Burgen, A. S. V. *Proc. R. Soc. B* **196**, 251–265 (1977).
17. Bacanari, D. P., Averett, D., Briggs, C. & Burchall, J. *Biochemistry* **16**, 3566–3572 (1977).
18. Cayley, P. J. *et al. Biochemistry* **18**, 3886–3895 (1979).
19. Gronenborn, A. *et al. Molec. Pharmac.* (in the press).

BOOK REVIEWS

Computers born of crisis

S.J. Goldsack

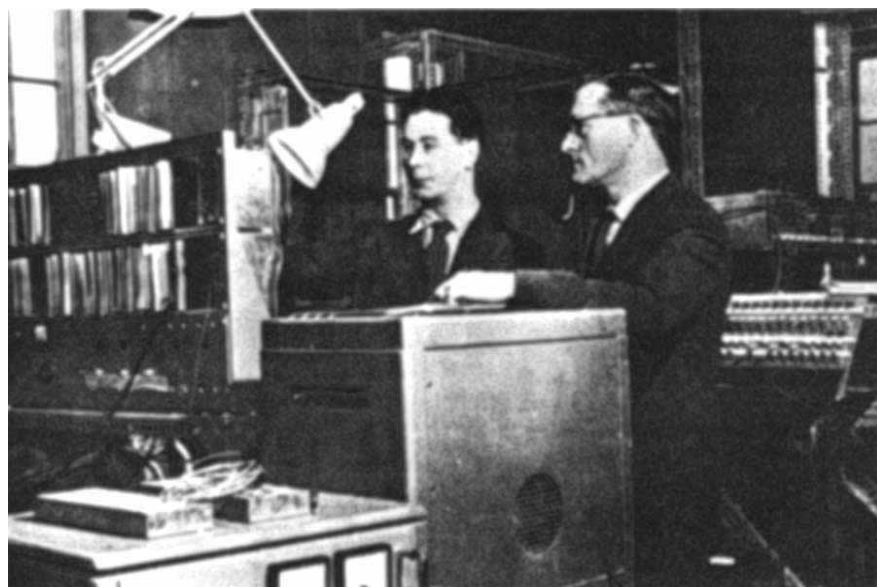
A History of Computing in the Twentieth Century. Edited by N. Metropolis *et al.* Pp.688. (Academic: 1980.) \$29.50, £16.60.
Early British Computers. By S. Lavington. Pp.139. (Manchester University Press: 1980.) Hbk £7.95; pbk £3.95.

SEPARATELY, these books would both be well worth reading by anyone with an interest in the origins of computing as we know it today. Together I found them fascinating. Both volumes claim to be of interest to the non-specialist reader, as well as to the technically aware. I would judge that this claim is reasonably justified; however, I think that an understanding of the technical background would certainly help.

The book by Metropolis *et al.* is the more important. In June 1976, a conference on the history of computing was held in Los Alamos. Its purpose was to bring together as many as possible of the pioneers who created the modern computer, to tell the story of their work. The timing was none too early; already important figures such as J. von Neumann, A. Turing and F.C. Williams, whose presence would have added greatly to the proceedings, had died. Their places were taken by colleagues who had shared in their work. Speakers described the early work on the design and use of computers in America, England, Germany, Russia and Japan. The book is essentially the written record of that conference.

Many of the basic concepts involved in computing were understood before the outbreak of war in 1939. However, the urgent need for tools of computation in connection with war work, and the great technical advances in electronics which were brought about by the efforts in radar and other fields, produced circumstances in which the construction of electronic computers could prosper.

It is commonplace to recognize how the threat of war in the 1930s, and the war effort that followed, fostered technical advances in many fields. The participants in these efforts experienced a uniquely exciting period in their lives. The intrinsic scientific interest of the work in which they were involved, the urgency with which the research was pursued, and, perhaps for the first time, a political commitment by the governments involved to ensure that the resources required were as far as possible made available, combined to create an atmosphere which had not before been



M.J. Lanigan and Tom Kilburn with a section of the transistorized MUSE computer in 1959. This prototype was a forerunner of the Ferranti ATLAS, inaugurated by Sir John Cockcroft in 1962, which incorporated several new concepts in computer design.

encountered. After 1945, those involved returned to the universities, or to newly established research centres, fired with enthusiasm to continue the work which they had started in the urgent circumstances of international crisis. As a student in the immediate post-war period, it was impossible not to sense the enthusiasm which our teachers brought to their work.

The authors of this book were not primarily writing memoirs of their experiences; yet the effect of the collection of accounts by all the parties participating in work relating to electronic computers about how progress was achieved in the different centres, and the interrelation between them, creates a strong impression of the vitality of the period.

It would not be realistic to single out particular papers as worthy of special notice, as the interest of the book lies in the way in which the contributions of the various authors combine to create a complete picture. Inevitably, however, there is special interest in those papers describing work done in conditions of secrecy during the war, and of which these accounts are among the first publicly available. They include, for example, the description of the Colossus machines, used to decipher German codes by the British Secret Service — oddly, this work was sponsored by the Foreign Office! I personally learned much of which I was

ignorant, or only dimly aware. I had failed to appreciate, for example, how widely the Williams storage tube was used on both sides of the Atlantic for a period before the invention of magnetic core stores.

The volume addresses the history of computing. Though hardware has pride of place, the developments in programming and programming languages are included. It would be impossible to appreciate the development of computer architecture, without understanding the need for better facilities for the programmer.

The date chosen, tacitly, to determine the end of history and the start of the modern era was rather arbitrary. Historical computers, it seems, used valves: transistors belong to the modern era. The computers of the second generation, which saw the development of operating systems and complex software, were not covered. Perhaps a future conference should address the history of programming and software; it is doubtful, however, whether it would reach the same level of interest for the general reader as is achieved in this book.

It is not surprising, perhaps, that Lavington's volume should appear at about the same time as the other. Lavington presented the work of the Manchester group under F.C. Williams at Los Alamos, and, no doubt, his book was inspired in part by the research involved in

From *Early British Computers*

preparing his contribution to the larger work. Though it has less scope and significance, it does include an account of the ATLAS Computer which was largely developed at Manchester and which introduced the concept of virtual storage in the early 1960s.

Lavington's early chapters, which attempt to give in a few pages a technical account of the early developments, are rather shaky. However, the style quickly settles down to an interesting description which reinforces and extends the material of the larger book. □

S.J. Goldsack is Professor in the Department of Computing at Imperial College, University of London.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Charles Babbage, 1792–1871, who devised the basic principles of modern computers.



Alan Turing, a leading figure in the early days of computer design. He died in 1954.

Mary Evans

Computing and optics in context

P.W. Hawkes

The Computer in Optical Research. Edited by B.R. Frieden. Pp.371. (Springer-Verlag: 1980.) DM 98, \$57.90.

SOME topics seem so naturally adapted to a multi-author series such as this that when the corresponding volume does appear it is hard to realize that it has no rivals. Frieden's collection is a good example: most of the material is available elsewhere, scattered throughout the literature though not necessarily that of optics, but how very much more useful and usable it becomes, gathered into a single volume, put in context and with a recognizable pattern imposed on it.

Five subject-areas are explored. First, R. Barakat describes the calculation of the various types of integrals that arise in diffraction theory; his aim is to draw attention to the methods that have been developed by numerical analysts, with which specialists in other fields are all too often unfamiliar. This is followed by a long and extremely important chapter by Frieden on probability and statistics. This contains a formal but highly readable account of those aspects of the subjects that Frieden believes to be valuable in optics. After arguing persuasively that such methods are likely to become all-pervasive in this field, he regrets that progress is slow since "most optical workers simply do not know the methods" even though "first applications in optics have been highly successful". He claims, quite justifiably, that his chapter is the "first exposition on probability and statistics written specifically from the

optical perspective and for the optical community".

The remainder of the book is less revolutionary though of considerable interest. A.K. Rigler and R.J. Pegis give a knowledgeable account of optimization methods, L. Mertz examines computers and optical astronomy and W.J. Dallas surveys the computer generation of holograms. Every chapter is filled with practical information, illustrated with numerous examples, and it is clear either that Frieden has submitted the chapters to thorough editing to ensure uniformity of tone or that he chose his contributors very skilfully. Furthermore, he has written an introductory chapter that adds substantially to the usefulness and readability of the remainder. In this, he writes what amounts to a long essay-review of the book, explaining informally how the individual chapters are organized and giving his own opinions of the place of particular approaches in optics. In addition he draws attention, sometimes anecdotally, to other topics that could have been included, or are at too preliminary a stage of development to appear in the main text; Fienup's modified version of the Gerchberg-Saxton algorithm is an example of this.

It is not often that the multi-author formula is entirely satisfactory but here it works to perfection. The best way of recommending this wholly excellent book is to say that I cannot understand how we have been managing without it. □

P.W. Hawkes is Maître de Recherches at the Laboratoire d'Optique Electronique du CNRS, Toulouse.

The double layer in electrochemistry

Andrew Hamnett

The Double Layer. Comprehensive Treatise on Electrochemistry, Vol.1. Edited by J. O'M. Bockris, B.E. Conway and E. Yeager. Pp.453. (Plenum: 1980.) \$49.50, £31.19.

THE Plenum Press have published a remarkable series of books on electrochemistry in recent years, and the *Comprehensive Treatise on Electrochemistry* is their latest and most ambitious venture. This first volume contains nine articles on aspects of the double layer written by such well-known electrochemists as Parsons, Bockris, Frumkin and Damaskin. Inevitably, these articles make rather varied intellectual demands on the reader, from the considerable in the case of Parsons' introductory article on thermodynamics, to the pedagogically straightforward

account by Reeves of the double layer without specific adsorption.

Parsons' account of the Gibbs adsorption isotherm is rightly given the task of introducing this volume. It is an essay of considerable erudition, though the development of the basic equations is very abstract. Some didactic edge is therefore lost in the early part of the essay and only recovered when the author considers some specific examples of electrode configuration.

The two articles by Reeves and by Habib and Bockris constitute the central part of the book. They are both clear and well written, though I would have preferred a more thorough and detailed account of the recent developments in double-layer theory described in Reeves' article. The essay by Bockris and Habib on specific ionic adsorption is a thorough account of a

highly controversial area; indeed, some of the flavour of this controversy is evident in the article, lending it at times an almost polemical tone. Equally clear is the discussion by Damaskin and Kazarimov on the adsorption of organic molecules onto electrodes; from the pedagogical point of view, this is the most successful article in the book.

Frumkin and co-authors present a detailed account of the concept of potential of zero charge. This topic has attracted several recent reviews and a book (by Frumkin); the present article is in fact a reprise of that book and will be welcomed therefore in the West. The importance of the concept of pzc is amply illustrated in the

present review and my only criticism is that there remain some problems in notation, both between this article and others in the volume, and indeed between text and figure captions.

The articles by Pleskov, Boguslavsky and Hunter are devoted to various aspects of the non-metal/electrolyte interface. Pleskov's essay reviews the double-layer on classical semiconductors and is an able summary of the famous book by Myamlin and Pleskov (*Electrochemistry of Semiconductors*; Plenum, 1967), but I was disappointed to find so few references to work of the past decade, especially given the explosion of interest that has occurred in this field. Boguslavsky has provided a

welcome if rather tentative account of the insulator/electrolyte interface and Hunter a brief but most useful description of colloids.

The present volume contains some essays of great value; in others the development is rather eclectic and the overall result is somewhat uneven. Nevertheless, the first volume of this series will undoubtedly become a standard work. Nothing quite so detailed has been attempted before in electrochemistry, and firmer editorial control should ensure that subsequent volumes will more than fulfil its promise. □

Andrew Hamnett is University Lecturer in Inorganic Chemistry and Fellow of St Catherine's College, Oxford.

Population genetics with age — not forgetting the genes

Michael Bulmer

Evolution in Age-Structured Populations. By Brian Charlesworth. Pp.300. (Cambridge University Press: 1980.) Hbk £18, \$44.50; pbk £5.95, \$13.50.

THIS is the first book in a new series called *Cambridge Studies in Mathematical Biology*, the aim of which is that "each book will be self-contained, covering necessary background theory, but without allowing the mathematical details to obscure biological relevance". The book under review admirably fulfils these aims.

The classical theory of population genetics assumes a life-history with discrete, non-overlapping generations in which all individuals in the population at a particular time are the same age. This model is appropriate for annual plants and insects, but not for a large number of species (most vertebrates, for example) with overlapping generations. In the latter situation, the population at a particular time consists of individuals of different ages with fecundities and survival probabilities dependent on their age, and a proper theory of its population genetics should take its age structure into account. In particular, it is clearly impossible to consider how natural selection moulds the way in which fecundity and survival depend on age without considering age structure, so that recent interest in the evolution of life-history strategies has stimulated theoreticians to extend classical population genetics theory to include age-structure as a variable. Brian Charlesworth reviews and summarizes this work, much of which is his own.

Five chapters develop the subject in a logical way. The first introduces the relevant demographic theory, and the next discusses population genetics in age-structured populations without selection (Hardy-Weinberg and linkage equilibrium, and genetic drift in finite populations); of particular value is the

description of the elegant work of Hajnal on the frequencies of consanguineous matings in an isolated population.

Chapters 3 and 4 form the core of the book and describe the effect of selection in an age-structured population. The author remarks that important work on this subject was published by Haldane and Norton in the late 1920s, but that further work was inhibited by the introduction by Fisher of his "Malthusian parameter" method in 1930, which apparently provided a simple and elegant answer to the problem. However, it has been recognized in the past ten years that this method is defective because a particular genotype cannot have a fixed Malthusian parameter when the genotypic composition of the population is changing under selection; there is no short-cut to the methodology introduced by Haldane and Norton.

Finally, Charlesworth discusses the evolution of life-history strategies, that is to say the evolution of senescence and of reproductive patterns. This is the most interesting area of application of the ideas developed in the book; the author's general conclusion is that

although qualitative agreement between theoretical expectation and observed differences in life-histories between and within species is often good, no studies of natural populations have really been capable of demonstrating close quantitative agreement between a real and an optimal life-history.

Few would disagree with this comment.

This is an important book which can be recommended to anyone with an interest in population genetics and evolutionary biology. The argument is presented clearly from first principles; the author does not belong to the hand-waving school of thought which believes that, when the going gets difficult, it is possible to do population genetics without genes. The mathematical level is reasonably straightforward, but it is at times difficult to see the

wood for the trees; a reader might do well to make a rapid first reading before studying the details.

Incidentally, it should be noted that Bernardelli is throughout mis-spelt "Bernadelli"; this is probably a copying error from a secondary source — a classical deletion mutation. □

Michael Bulmer is a Lecturer in Biomathematics at Oxford University.

SCIENTIFIC BOOKSHOP

H.K. LEWIS can supply works in all branches of Pure and Applied Science. Catalogues on request. Please state interests.

SCIENTIFIC LENDING LIBRARY

Annual Subscription from £7.50. (Available in U.K. only)

Reduced rates for multiple subscriptions.

Prospectus post free on request.

Quarterly List of New Books and new editions added to the Library sent post free to subscribers regularly.

H.K. LEWIS & Co. Ltd.

136 GOWER STREET,
LONDON, WC1E 6BS

Telephone: 01-387 4282
Telegrams: "Publicavit,
London, WC1E 6BS."

Circle No.26 on Reader Enquiry Card.

ISSN-0275-4207

**THE BIOTECHNOLOGY
NEWS SOURCE™**

BioEngineering News -
the *only* biweekly
newsletter covering the
genetic-engineering
recombinant-DNA,
monoclonal-antibody,
and cell-culture
industries.

24 issues — \$236.00 (U.S.)
(\$195.00 in U.S. & Canada)

Subscriptions sent airmail.

Write: **BioEngineering News**, Suite
304, 109 Minna St., San Francisco,
CA 94105. Tel: (415) 332-0508.
Telex: 171463-HQLARK.

Special Reports Available: (1) Genetic Engineering Industry, (2) Hybridoma & Cell Fusion, (3) Automated DNA Synthesizers. Price: \$29.95 (U.S. Funds).

NOTE - For prepaid subscriptions select any two (2) reports as a bonus.

Circle No.04 on Reader Enquiry Card.

**MICROCOMPUTERS
and PHYSIOLOGICAL
SIMULATION**

James E. Randall, *Indiana University
School of Medicine, Bloomington*
L. S. Arthur C. Gavton, *University of
Mississippi School of Medicine, Jackson*

"Here is where Dr. Randall's book is invaluable: the background information on microprocessors, combined with specific examples of biological data simulated with various hardware and software configurations, should allow any life-science experimenter to progress rapidly from the initial idea to a working simulation model."

BYTE

Contents:

Introduction. Microcomputer Components. Operating Systems and Programming Languages. Hardware Enhancements for Simulation. Representative Microcomputers. Compartmental Kinetics: A First Example. The Glucose Tolerance Test. Cardiovascular System Mechanics. Arterial Pulse Pressure. Vectorcardiography and the Limb Leads. Distortion of Waveforms. Axon Action Potentials. Cardiac Action Potentials. Formatting Student Exercises. Index.

1980. 250 pp., illus. Paper 06128 \$14.50

Brochure Available from Publisher.
Price is subject to change without notice.
Price higher outside U.S.A.

**Addison-Wesley**

Advanced Book Program,
World Science Division
Reading, Mass. 01867, U.S.A.

London, Amsterdam, Don Mills, Ontario, Sydney, Tokyo

Circle No.11 on Reader Enquiry Card.

Some worthy publicity for elastin

A. Serafini-Fracassini

Biology and Pathology of Elastic Tissues. Frontiers of Matrix Biology, Vol.8. Edited by A.M. Robert and L. Robert. Pp.230. (S. Karger: 1980.) \$74.25, SwFr.124.

THIS book attempts to expound, for the benefit of graduate students and professional investigators, various areas of scientific research which are contained within the subject of molecular biology and pathology of elastin. This is undoubtedly a topic of considerable importance to biochemical and medical workers, and one which has attracted a large amount of research over the past two decades. As unambiguous interpretations of functional properties of the protein still prove elusive, it is of some value to have, brought together in one volume, a collection of critical evaluations of some recent findings. The earlier contributions in the book deal with structural and conformational aspects of the protein, while the others discuss its implication in the inflammatory response and in some tissue disorders. The role of elastases in these pathological conditions is adequately outlined. However, in this context, one would like to have seen included an up-to-date account of the biosynthesis of elastin, with particular reference to the nature of the soluble biosynthetic intermediate and

to the biogenesis of the cross-links. Such a chapter would have been of great value to an outsider to the subject in appreciating the significance of some structural features of the protein and in understanding the pathogenesis of specific disorders. Other aspects which with advantage could have been covered more fully are the physico-chemical properties of elastin, and the thermodynamics of its deformation to balance Dr Urry's excellent contribution on synthetic repeat polypeptides.

It is difficult to make a unified judgement on this collection of papers because the style of the contributions is so varied. Some offer an accurate review of the relevant scientific research, while others mainly outline the results obtained in the authors' laboratories, occasionally with the exclusion of some lines of research that might have merited discussion. This is also reflected by the variable number of references cited at the end of each chapter.

The book is well produced, adequately illustrated and as a whole very readable. Elastin has not yet enjoyed its share of the general escalation in scientific publications, and I consider this a worthy collection, one which could be read with profit in its entirety by non-specialists. □

A. Serafini-Fracassini is Professor of Biochemistry at the University of St Andrews.

Compleat diquat and paraquat

R.J. Hance

The Bipyridinium Herbicides. By L.A. Summers. Pp.449. (Academic: 1980.) £33, \$79.50.

ONLY two bipyridinium herbicides, diquat and paraquat, are in widespread use, so it is interesting to speculate why they have stimulated a book of this size when larger groups of herbicides have not. Perhaps there are two reasons. The first is the enthusiasm of Dr Summers for his subject. The second is that, in comparison with other herbicides, the peripheral synthetic chemistry of these compounds is challenging and their toxicology unusual. Indeed, these topics occupy almost half the book. The remainder covers the chemistry of the respective diquaternary salts, structure-activity relationships, environmental fate, occurrence and analysis of residues, herbicidal use and modes of action.

With 2,500 references cited in little over 400 pages the text is hard going in places but, by intelligent use of tables, the author has done his best to avoid producing a catalogue. The only reasonable criticism that might be made is that not all readers will follow the logic underlying the sequence in which subjects are presented.

For instance, the chapters dealing with fate in biological systems and with residues are separated from that on toxicology and environmental effects which in many respects is closely related. The layout within chapters also occasionally seems strange. There is a good example on p.394, where a paragraph describing the most effective ways of treating paraquat poisoning is placed in the middle of a section rather than at the end where it would have more impact. However, the book is sensibly indexed so the reader should have little difficulty in finding specific information.

Nowadays, books including a range of specialist subjects tend to be multi-authored. For one man to tackle such a task successfully is an impressive feat of scholarship and Dr Summers deserves warm congratulation. His book undoubtedly will become a standard reference. It shows, incidentally, that there remains a place for books of this type alongside the computer-based information retrieval systems that seem to threaten their viability. □

R. J. Hance is leader of the Herbicide Group at the Agricultural Research Council Weed Research Organization, Yarnton, near Oxford.